# BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

## IMAGES ARE BEST AVAILABLE COPY.

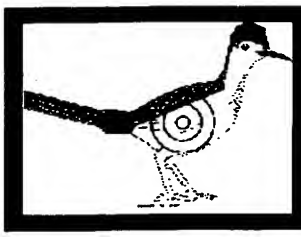
**As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.**

FIG. 1



File Edit View Go Communicator Help

Bookmarks Location: http: //victoria. inpharmatica.co.uk/~volker/BPD3target.html



Target Mining Interface

inpharmatica

## Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): 1lfa and chain (e.g. B): a

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

## Select Database

Release: DEVF9 = BPD3 ☐

## Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

Matches detected during the first 20 forward iterations ☐

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:

☐Homo sapiens | ☐Rattus norvecicus (Rat) | ☐Mus musculus (Mouse) | ☐Danio rerio (Zebra fish)

100%

FIG. 2A



File Edit View Go Communicator Help

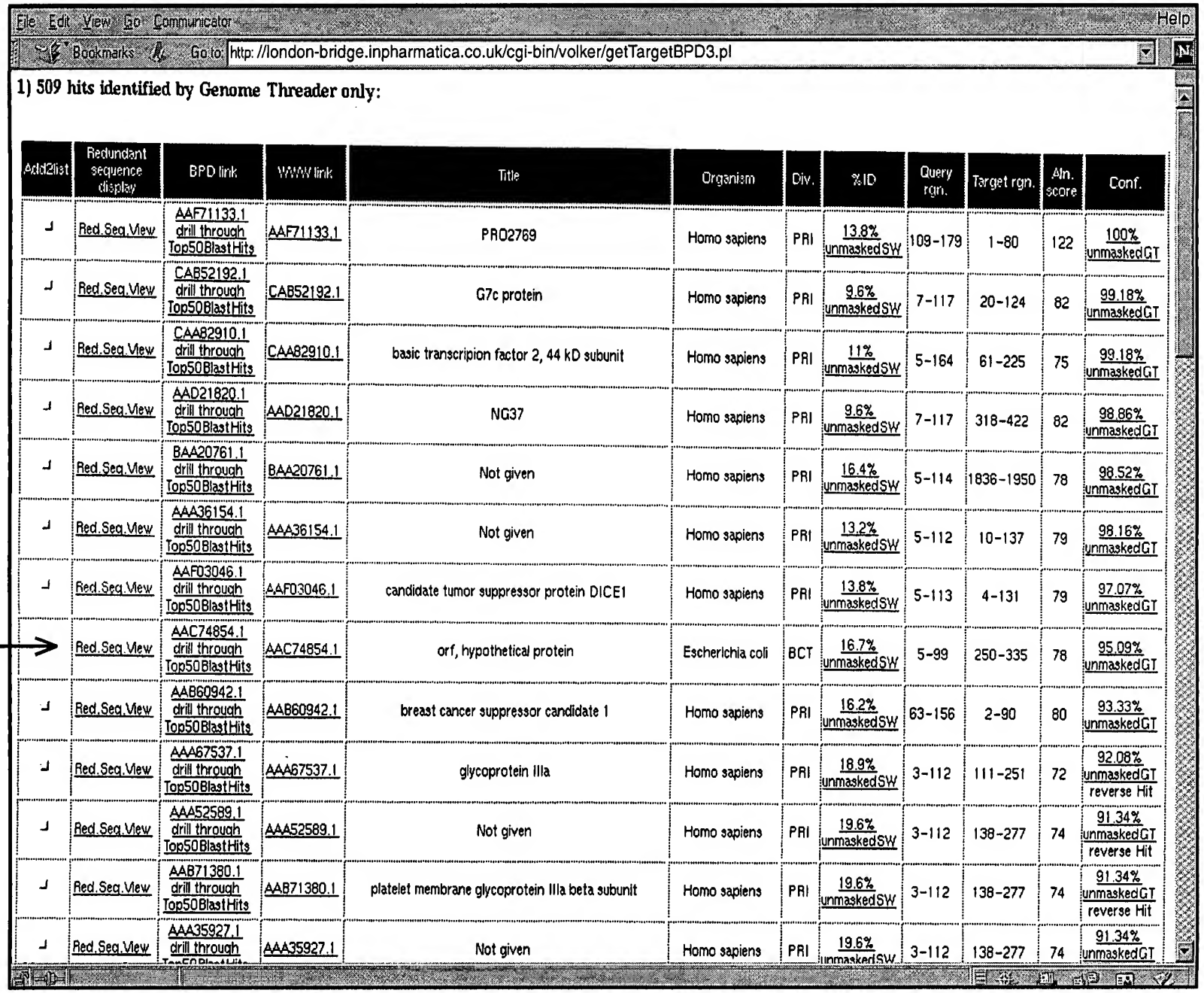
Bookmarks Location: http: //London-bridge. inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl

**2) 84 additional hits identified by both, Genome Threader and PSI-BLAST:**

**Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!**

| Add2list | BPD links | WWW link | Title | Organism | Div. | %ID (GT,PSI) | Query rgn. (GT,PSI) | Target rgn. (GT,PSI) | Aln. score (GT) | Conf. (GT) | 1st Iter. (PSI) | Best Iter. (PSI) | Be E-v (P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☐ | AAA59544.1 drill through Top50BlastHits Red.Seq.View | AAA59544.1 | Not given | Homo sapiens | PRI | 33.7%, 33% unmaskedSW | 4-183, 4-183 | 150-336, 150-336 | 449 | 100% unmaskedGT | 1 | 2 | 3E- |
| ☐ | AAB24821.1 drill through Top50BlastHits Red.Seq.View | AAB24821.1 | leukocyte integrin alpha chain | Homo sapiens | PRI | 33.7%, 33% unmaskedSW | 4-183, 4-183 | 150-336, 150-336 | 449 | 100% unmaskedGT | 1 | 2 | 3E- |
| ☐ | Q99715 drill through Top50BlastHits Red.Seq.View | Q99715 | COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR. | Homo sapiens (Human). | PRI | 28.8%, 26% unmaskedSW | 4-181, 2-174 | 140-318, 2321-2495 | 440 | 100% unmaskedGT | 1 | 3 | 2E- |
| ☐ | P20701 drill through Top50BlastHits Red.Seq.View | P20701 | LEUKOCYTE ADHESION GLYCOPROTEIN LFA-1 ALPHA CHAIN PRECURSOR (LEUKOCYTE FUNCTION ASSOCIATED MOLECULE 1, ALPHA CHAIN) (CD11A) (INTEGRIN ALPHA-L). | Homo sapiens (Human). | PRI | 100%, 100% unmaskedSW | 1-183, 1-183 | 153-335, 153-335 | 423 | 100% unmaskedGT | 1 | 1 | 1E- |
| ☐ | AAC31672.1 drill through Top50BlastHits Red.Seq.View | AAC31672.1 | leukocyte function-associated molecule-1 alpha subunit | Homo sapiens | PRI | 99.5%, 99% unmaskedSW | 1-183, 1-183 | 153-335, 153-335 | 423 | 100% unmaskedGT | 1 | 1 | 1E- |
| ☐ | CAA72402.1 drill through Top50BlastHits Red.Seq.View | CAA72402.1 | collagen type XIV | Homo sapiens | PRI | 29.1%, 29% unmaskedSW | 2-180, 2-180 | 5-185, 5-185 | 422 | 100% unmaskedGT | 1 | 2 | 2E- |
| ☐ | AAB38702.1 drill through Top50BlastHits Red.Seq.View | AAB38702.1 | cartilage matrix protein | Homo sapiens | PRI | 31.7%, 27% unmaskedSW | 4-183, 2-182 | 275-455, 39-223 | 413 | 100% unmaskedGT | 1 | 2 | 2E- |
| ☐ | CAB70853.1 drill through Top50BlastHits Red.Seq.View | CAB70853.1 | hypothetical protein | Homo sapiens | PRI | 28%, 28% unmaskedSW | 1-180, 1-183 | 437-620, 437-624 | 406 | 100% unmaskedGT | 1 | 2 | 1E- |
| ☐ | CAA27972.1 drill through Top50BlastHits Red.Seq.View | CAA27972.1 | Not given | Homo sapiens | PRI | 20.5%, 20% unmaskedSW | 3-183, 2-181 | 1497-1673, 1689-1873 | 405 | 100% unmaskedGT | 2 | 3 | 1E- |
| ☐ | AAB59512.1 drill through Top50BlastHits Red.Seq.View | AAB59512.1 | Not given | Homo sapiens | PRI | 20.5%, 20% unmaskedSW | 3-183, 2-181 | 758-934, 950-1134 | 405 | 100% unmaskedGT | 2 | 3 | 1E- |
| ☐ | CAA07569.1 drill through Top50BlastHits | CAA07569.1 | matrilin-4 | Homo sapiens | PRI | 28.1%, 25% unmaskedSW | 1-183, 1-183 | 342-528, 31-217 | 403 | 100% unmaskedGT | 1 | 2 | 2E- |

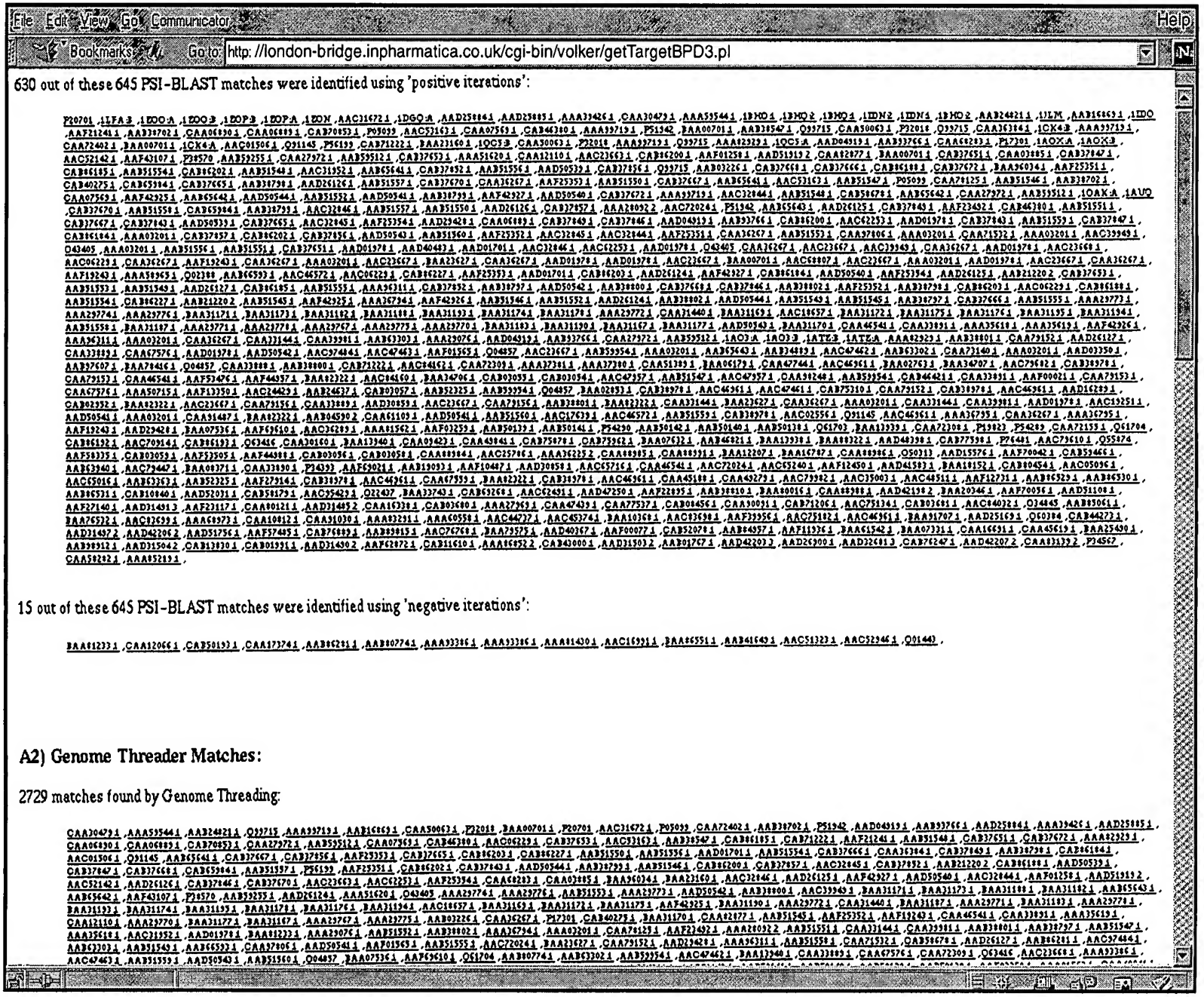
FIG. 2B




**1) 509 hits identified by Genome Threader only:**

| Add2list | Redundant sequence display | BPD link | WWW link | Title | Organism | Div. | %ID | Query rgn. | Target rgn. | Aln. score | Conf. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ⌟ | Red.Seq.View | AAF71133.1 drill through Top50BlastHits | AAF71133.1 | PRO2769 | Homo sapiens | PRI | 13.8% unmaskedSW | 109-179 | 1-80 | 122 | 100% unmaskedGT |
| ⌟ | Red.Seq.View | CAB52192.1 drill through Top50BlastHits | CAB52192.1 | G7c protein | Homo sapiens | PRI | 9.6% unmaskedSW | 7-117 | 20-124 | 82 | 99.18% unmaskedGT |
| ⌟ | Red.Seq.View | CAA82910.1 drill through Top50BlastHits | CAA82910.1 | basic transcripion factor 2, 44 kD subunit | Homo sapiens | PRI | 11% unmaskedSW | 5-164 | 61-225 | 75 | 99.18% unmaskedGT |
| ⌟ | Red.Seq.View | AAD21820.1 drill through Top50BlastHits | AAD21820.1 | NG37 | Homo sapiens | PRI | 9.6% unmaskedSW | 7-117 | 318-422 | 82 | 98.86% unmaskedGT |
| ⌟ | Red.Seq.View | BAA20761.1 drill through Top50BlastHits | BAA20761.1 | Not given | Homo sapiens | PRI | 16.4% unmaskedSW | 5-114 | 1836-1950 | 78 | 98.52% unmaskedGT |
| ⌟ | Red.Seq.View | AAA36154.1 drill through Top50BlastHits | AAA36154.1 | Not given | Homo sapiens | PRI | 13.2% unmaskedSW | 5-112 | 10-137 | 79 | 98.16% unmaskedGT |
| ⌟ | Red.Seq.View | AAF03046.1 drill through Top50BlastHits | AAF03046.1 | candidate tumor suppressor protein DICE1 | Homo sapiens | PRI | 13.8% unmaskedSW | 5-113 | 4-131 | 79 | 97.07% unmaskedGT |
| → | Red.Seq.View | AAC74854.1 drill through Top50BlastHits | AAC74854.1 | orf, hypothetical protein | Escherichia coli | BCT | 16.7% unmaskedSW | 5-99 | 250-335 | 78 | 95.09% unmaskedGT |
| ⌟ | Red.Seq.View | AAB60942.1 drill through Top50BlastHits | AAB60942.1 | breast cancer suppressor candidate 1 | Homo sapiens | PRI | 16.2% unmaskedSW | 63-156 | 2-90 | 80 | 93.33% unmaskedGT |
| ⌟ | Red.Seq.View | AAA67537.1 drill through Top50BlastHits | AAA67537.1 | glycoprotein IIIa | Homo sapiens | PRI | 18.9% unmaskedSW | 3-112 | 111-251 | 72 | 92.08% unmaskedGT reverse Hit |
| ⌟ | Red.Seq.View | AAA52589.1 drill through Top50BlastHits | AAA52589.1 | Not given | Homo sapiens | PRI | 19.6% unmaskedSW | 3-112 | 138-277 | 74 | 91.34% unmaskedGT reverse Hit |
| ⌟ | Red.Seq.View | AAB71380.1 drill through Top50BlastHits | AAB71380.1 | platelet membrane glycoprotein IIIa beta subunit | Homo sapiens | PRI | 19.6% unmaskedSW | 3-112 | 138-277 | 74 | 91.34% unmaskedGT reverse Hit |
| ⌟ | Red.Seq.View | AAA35927.1 drill through | AAA35927.1 | Not given | Homo sapiens | PRI | 19.6% unmaskedSW | 3-112 | 138-277 | 74 | 91.34% unmaskedGT |

FIG. 2C




630 out of these 645 PSI-BLAST matches were identified using 'positive iterations':

P20701 ,1LFA:B ,1EDO:A ,1EDO:B ,1EOP:B ,1EOP:A ,1EON ,AAC31672.1 ,1DGQ:A ,AAD25884.1 ,AAD25885.1 ,AAA33426.1 ,CAA30479.1 ,AAA59544.1 ,1BMO:A ,1BMO:B ,1BMQ:A ,1IDN2 ,1IDN4 ,1BMO:B ,AAB24421.1 ,1ULM ,AAB56869.1 ,1IDO ,AAF21241.1 ,AAB38702.1 ,CAA06830.1 ,CAA06889.1 ,CAB70853.1 ,P05099 ,AAC52163.1 ,CAA07569.1 ,CAB46380.1 ,AAA59719.1 ,P51542 ,BAA00701.1 ,AAB38547.1 ,Q99715 ,CAA50063.1 ,P32018 ,Q99715 ,CAA36384.1 ,1CK4:B ,AAA59719.1 ,CAA72402.1 ,BAA00701.1 ,1CK4:A ,AAC01506.1 ,Q91145 ,P56199 ,CAB71222.1 ,BAA22160.1 ,1QCS:B ,CAA50063.1 ,P32018 ,AAA59719.1 ,Q99715 ,AAA42329.1 ,1QCS:A ,AAD04919.1 ,AAB39766.1 ,CAA68283.1 ,P17201 ,1AOX:A ,1AOX:B ,AAC52142.1 ,AAF43107.1 ,P39570 ,AAB59255.1 ,CAA27972.1 ,AAB59512.1 ,CAB37653.1 ,AAA51620.1 ,CAA12110.1 ,AAC23663.1 ,CAB46200.1 ,AAF01258.1 ,AAD51919.2 ,CAA82877.1 ,BAA00701.1 ,CAB37651.1 ,CAA03885.1 ,CAB37847.1 ,CAB46185.1 ,AAB51594.1 ,CAB46202.1 ,AAB51548.1 ,AAC31952.1 ,AAB65644.1 ,CAB37852.1 ,AAB51556.1 ,AAD50532.1 ,CAB37856.1 ,Q99715 ,AAB03226.1 ,CAB37668.1 ,CAB37666.1 ,CAB46188.1 ,CAB37672.1 ,BAA96034.1 ,AAF25351.1 ,CAB40275.1 ,CAB65944.1 ,CAB37665.1 ,AAB38798.1 ,AAD26126.1 ,AAB51557.1 ,CAB37670.1 ,CAA36267.1 ,AAF25353.1 ,AAB51550.1 ,CAB37667.1 ,AAB65641.1 ,AAC53163.1 ,AAB51547.1 ,P05099 ,CAA78125.1 ,AAB51546.1 ,AAB38702.1 ,CAA07569.1 ,AAF42925.1 ,AAB65642.1 ,AAD50544.1 ,AAB51552.1 ,AAD50541.1 ,AAB38799.1 ,AAF42927.1 ,AAD50540.1 ,CAB37672.1 ,AAA59719.1 ,AAC32844.1 ,AAB51548.1 ,CAB56678.1 ,AAB65642.1 ,CAA27972.1 ,AAB59512.1 ,1OAK:A ,1AUQ ,CAB37670.1 ,AAB51558.1 ,CAB65944.1 ,AAB38799.1 ,AAC32846.1 ,AAB51557.1 ,AAB51550.1 ,AAD26126.1 ,CAB37857.1 ,AAA28092.2 ,AAC72024.1 ,P51542 ,AAB65643.1 ,AAD26125.1 ,CAB37843.1 ,AAF23452.1 ,CAB46380.1 ,AAB51551.1 ,CAB37667.1 ,CAB37843.1 ,AAD50539.1 ,CAB37665.1 ,AAC32845.1 ,AAF25354.1 ,AAD29428.1 ,CAA06889.1 ,CAB37849.1 ,CAB37846.1 ,AAD04919.1 ,AAB39766.1 ,CAB46200.1 ,AAC62253.1 ,AAD01978.1 ,CAB37843.1 ,AAB51559.1 ,CAB37847.1 ,CAB46184.1 ,AAA03201.1 ,CAB37857.1 ,CAB46202.1 ,CAB37856.1 ,AAD50543.1 ,AAB51560.1 ,AAF25352.1 ,AAC32845.1 ,AAC32844.1 ,AAF25351.1 ,CAA36267.1 ,AAB51553.1 ,CAA97406.1 ,AAA03201.1 ,CAA71522.1 ,AAA03201.1 ,AAC39945.1 ,O43405 ,AAA03201.1 ,AAB51556.1 ,AAB51551.1 ,CAB37851.1 ,AAD01978.1 ,AAD40483.1 ,AAD01701.1 ,AAC32846.1 ,AAC62253.1 ,AAD01978.1 ,O43405 ,CAA36267.1 ,CAA36267.1 ,AAC39945.1 ,CAA36267.1 ,AAD01978.1 ,AAC23668.1 ,AAC06229.1 ,CAA36267.1 ,AAF19243.1 ,CAA36267.1 ,AAA03201.1 ,AAC23667.1 ,BAA23627.1 ,CAA36267.1 ,AAD01978.1 ,AAD01978.1 ,AAC23667.1 ,BAA00701.1 ,AAC68807.1 ,AAC23667.1 ,AAA03201.1 ,AAD01978.1 ,AAC23667.1 ,CAA36267.1 ,AAF19243.1 ,AAA58965.1 ,Q02388 ,AAB66593.1 ,AAC46572.1 ,AAC06229.1 ,CAB46227.1 ,AAF25353.1 ,AAD01701.1 ,CAB46203.1 ,AAD26124.1 ,AAF42927.1 ,CAB46184.1 ,AAD50540.1 ,AAF25354.1 ,AAD26125.1 ,AAB21220.2 ,CAB37653.1 ,AAB51553.1 ,AAB51549.1 ,AAD26127.1 ,CAB46185.1 ,AAB51555.1 ,AAA96311.1 ,CAB37852.1 ,AAB38797.1 ,AAD50542.1 ,AAB38800.1 ,CAB37668.1 ,CAB37846.1 ,AAB38802.1 ,AAF25352.1 ,AAB38798.1 ,CAB46203.1 ,AAC06229.1 ,CAB46188.1 ,AAB51554.1 ,CAB46227.1 ,AAB21220.2 ,AAB51545.1 ,AAF42925.1 ,AAA26794.1 ,AAF42926.1 ,AAB51546.1 ,AAB51552.1 ,AAD26124.1 ,AAB38802.1 ,AAD50544.1 ,AAB51549.1 ,AAB38797.1 ,CAB46226.1 ,AAB51555.1 ,AAA26773.1 ,AAA26794.1 ,AAA26776.1 ,BAA31173.1 ,BAA31172.1 ,BAA31182.1 ,BAA31188.1 ,BAA31193.1 ,BAA31174.1 ,BAA31187.1 ,BAA31175.1 ,BAA31176.1 ,BAA31195.1 ,BAA31194.1 ,AAB51558.1 ,BAA31197.1 ,AAA26771.1 ,AAA26770.1 ,AAA26767.1 ,AAA26779.1 ,AAA26770.1 ,BAA31183.1 ,BAA31190.1 ,BAA31167.1 ,BAA31177.1 ,AAD50543.1 ,BAA31170.1 ,CAA46541.1 ,CAA28951.1 ,AAA35618.1 ,AAA35619.1 ,AAF42926.1 ,AAA26211.1 ,AAA03201.1 ,CAA36267.1 ,CAA23544.1 ,CAA35991.1 ,AAB63303.1 ,AAA26776.1 ,AAD04913.1 ,AAB53766.1 ,CAA26772.1 ,AAB55512.1 ,1AO3:A ,1AO3:B ,1ATE:B ,1ATE:A ,AAA82923.1 ,AAB38801.1 ,CAA79152.1 ,AAD26127.1 ,CAA23883.1 ,CAA67576.1 ,AAD01978.1 ,AAD50542.1 ,AAC97484.1 ,AAC47463.1 ,AAF01565.1 ,Q04857 ,AAC23667.1 ,AAB59956.1 ,AAA03201.1 ,AAB65643.1 ,AAB34819.1 ,AAC47462.1 ,AAB63302.1 ,CAA72140.1 ,AAA03201.1 ,AAD03350.1 ,AAB37607.1 ,BAA78416.1 ,Q04857 ,CAA23883.1 ,AAB38800.1 ,CAB72222.1 ,AAC68452.1 ,CAA72309.1 ,AAA27381.1 ,AAA27380.1 ,CAA51389.1 ,BAA01791.1 ,CAA47744.1 ,AAC46561.1 ,BAA02763.1 ,BAA36707.1 ,AAC79582.1 ,CAB38978.1 ,CAA79152.1 ,CAA46541.1 ,AAF51476.1 ,AAF44957.1 ,BAA82222.1 ,AAC68460.1 ,BAA36706.1 ,CAB03054.1 ,AAC47357.1 ,AAB51547.1 ,AAC47457.1 ,CAA82548.1 ,AAB59564.1 ,CAB46421.1 ,CAA23681.1 ,AAF00021.1 ,CAA79153.1 ,CAA67576.1 ,AAA50715.1 ,AAF12350.1 ,AAC24429.1 ,AAB24437.1 ,CAB03057.1 ,AAB52325.1 ,AAB59564.1 ,Q04857 ,BAA02853.1 ,CAB38978.1 ,AAC46561.1 ,AAC47461.1 ,CAB75310.1 ,CAA79152.1 ,CAB38978.1 ,AAC46561.1 ,AAD16289.1 ,CAB02952.1 ,BAA82222.1 ,AAC23667.1 ,CAA79156.1 ,AAD30859.1 ,AAC23667.1 ,CAA79156.1 ,AAB38801.1 ,BAA82222.1 ,CAA23544.1 ,BAA22627.1 ,CAA36267.1 ,AAA03201.1 ,CAA23544.1 ,CAA39981.1 ,AAD01978.1 ,AAC15251.1 ,AAD50541.1 ,AAA03201.1 ,CAA51487.1 ,BAA82222.1 ,AAB04502.1 ,CAA61103.1 ,AAD50541.1 ,AAB51560.1 ,AAC17639.1 ,AAC46572.1 ,AAB51559.1 ,CAB38978.1 ,AAC02556.1 ,Q91145 ,AAC46561.1 ,AAA36795.1 ,CAA36267.1 ,AAA36795.1 ,AAF19243.1 ,AAD29428.1 ,BAA07536.1 ,AAF65610.1 ,AAC62889.1 ,AAA51562.1 ,AAF02259.1 ,AAB50139.1 ,AAB50141.1 ,P54290 ,AAB50142.1 ,AAB50140.1 ,AAB50138.1 ,Q61703 ,BAA13939.1 ,CAA72308.1 ,P19823 ,P54289 ,CAA72355.1 ,Q61704 ,CAB65192.1 ,AAC70914.1 ,CAB65193.1 ,Q63416 ,CAA20160.1 ,BAA13940.1 ,CAA09423.1 ,CAA43941.1 ,CAB75078.1 ,CAB75562.1 ,BAA07632.1 ,AAB46821.1 ,BAA13938.1 ,BAA80322.1 ,AAD48398.1 ,CAB77598.1 ,P75441 ,AAC75610.1 ,Q55526 ,AAF58335.1 ,CAB03059.1 ,AAF53505.1 ,AAF44988.1 ,CAB03056.1 ,CAB03058.1 ,CAA88984.1 ,AAC15796.1 ,AAA26225.2 ,CAA88985.1 ,CAA88983.1 ,BAA12207.1 ,BAA16787.1 ,CAA88986.1 ,Q50311 ,AAD15576.1 ,AAF70042.1 ,CAB52466.1 ,AAB63940.1 ,AAC72447.1 ,BAA00371.1 ,CAA33490.1 ,P14393 ,AAF69021.1 ,AAB19093.1 ,AAF10447.1 ,AAD30858.1 ,AAC57516.1 ,CAA46541.1 ,AAC72024.1 ,AAC62400.1 ,AAF12450.1 ,AAD45583.1 ,BAA18152.1 ,CAB80454.1 ,AAC05096.1 ,AAC65016.1 ,AAB63263.1 ,AAB52325.1 ,AAF27914.1 ,CAB38978.1 ,AAC46561.1 ,AAC47955.1 ,AAC46561.1 ,CAA46588.1 ,CAA46279.1 ,AAC79982.1 ,AAC35003.1 ,AAC46561.1 ,AAF12731.1 ,AAB96329.1 ,AAB96530.1 ,AAB96531.1 ,CAB10840.1 ,AAD52031.1 ,CAB61759.1 ,AAC35629.1 ,Q22407 ,BAA33743.1 ,CAB63268.1 ,AAC62491.1 ,AAD47250.1 ,AAF22895.1 ,AAB38810.1 ,BAA80016.1 ,CAA88988.1 ,AAD42158.2 ,BAA20346.1 ,AAF70056.1 ,AAD51108.1 ,AAF27140.1 ,AAD31493.1 ,AAF23117.1 ,CAA80121.1 ,AAD31495.2 ,CAA16328.1 ,CAB03080.1 ,AAA27969.1 ,CAA47439.1 ,CAA77537.1 ,CAB04656.1 ,CAA30091.1 ,CAB75206.1 ,AAC75134.1 ,CAB03681.1 ,AAC64022.1 ,O34445 ,AAB85061.1 ,BAA76522.1 ,AAC12699.1 ,AAA69373.1 ,CAA10812.1 ,CAA31030.1 ,AAA82291.1 ,AAA60558.1 ,AAC44737.1 ,AAC45374.1 ,BAA10369.1 ,AAC83698.1 ,AAF32956.1 ,AAC75182.1 ,AAC46361.1 ,BAA91707.1 ,AAD25169.1 ,Q50384 ,CAB44273.1 ,AAD31497.2 ,AAD42206.2 ,AAD51756.1 ,AAF57485.1 ,CAB76889.1 ,AAB89815.1 ,AAC76769.1 ,BAA79575.1 ,AAD40367.1 ,AAF00077.1 ,AAB52079.1 ,AAB46597.1 ,AAF11536.1 ,BAA61542.1 ,BAA07331.1 ,CAA16691.1 ,CAA45619.1 ,BAA25490.1 ,AAB98912.1 ,AAD31504.2 ,CAB13930.1 ,CAB01991.1 ,AAD31490.2 ,AAF62872.1 ,CAB11610.1 ,AAA86852.2 ,CAB43000.1 ,AAD31503.2 ,AAB01767.1 ,AAD42203.2 ,AAD25900.1 ,AAD32681.1 ,CAB76247.1 ,AAD42207.2 ,CAA81392.2 ,P34567 ,CAA58282.1 ,AAA85283.1 ,

15 out of these 645 PSI-BLAST matches were identified using 'negative iterations':

BAA81233.1 ,CAA12066.1 ,CAB50193.1 ,CAA17374.1 ,AAB62981.1 ,AAB07746.1 ,AAA93386.1 ,AAA93385.1 ,AAA81430.1 ,AAC16991.1 ,BAA65551.1 ,AAB61643.1 ,AAC51323.1 ,AAC52946.1 ,Q01443 ,

## A2) Genome Threader Matches:

2729 matches found by Genome Threading:

CAA30479.1 ,AAA59544.1 ,AAB24421.1 ,Q99715 ,AAA59719.1 ,AAB16869.1 ,CAA50063.1 ,P32018 ,BAA00701.1 ,P20701 ,AAC31672.1 ,P05099 ,CAA72402.1 ,AAB38702.1 ,P51542 ,AAD04919.1 ,AAB39766.1 ,AAD25884.1 ,AAA33426.1 ,AAD25885.1 ,CAA06830.1 ,CAA06889.1 ,CAB70853.1 ,CAA27972.1 ,AAB59512.1 ,CAA07569.1 ,CAB46380.1 ,AAC62253.1 ,CAB37653.1 ,AAC53163.1 ,AAB38547.1 ,CAB46185.1 ,CAB71222.1 ,AAF25241.1 ,AAB51548.1 ,CAB37651.1 ,CAB37672.1 ,AAA82329.1 ,AAC01506.1 ,Q91145 ,AAB65641.1 ,CAB37667.1 ,CAB37856.1 ,AAF25353.1 ,CAB37665.1 ,CAB46203.1 ,CAB46227.1 ,AAB51550.1 ,AAB51556.1 ,AAD01701.1 ,AAB51554.1 ,CAB37666.1 ,CAA36384.1 ,CAB37849.1 ,AAB38798.1 ,CAB46184.1 ,CAB37847.1 ,CAB37668.1 ,CAB65944.1 ,AAB51557.1 ,P56199 ,AAF25351.1 ,CAB46202.1 ,CAB37843.1 ,AAD50544.1 ,AAB38799.1 ,AAB51546.1 ,CAB46200.1 ,CAB37857.1 ,AAC32845.1 ,CAB37852.1 ,AAB21220.2 ,CAB37670.1 ,AAD50539.1 ,AAC52142.1 ,AAD26126.1 ,CAB37846.1 ,CAB37670.1 ,AAC23663.1 ,AAC62253.1 ,AAF25354.1 ,CAA68283.1 ,CAA03885.1 ,BAA96034.1 ,BAA23160.1 ,AAC32846.1 ,AAD26125.1 ,AAF42927.1 ,AAD50540.1 ,AAC32844.1 ,AAF01258.1 ,AAD51919.2 ,AAB65642.1 ,AAF43107.1 ,P39570 ,AAB59255.1 ,AAD26124.1 ,AAA51620.1 ,O43405 ,AAA29776.1 ,AAA29775.1 ,AAB51553.1 ,AAA29771.1 ,AAD50542.1 ,AAB38800.1 ,AAC39945.1 ,BAA31171.1 ,BAA31173.1 ,BAA31188.1 ,BAA31182.1 ,AAB65643.1 ,BAA31193.1 ,BAA31174.1 ,BAA31195.1 ,BAA31178.1 ,BAA31176.1 ,BAA31194.1 ,AAC46572.1 ,AAF42925.1 ,BAA31190.1 ,AAA29772.1 ,CAA21440.1 ,BAA31187.1 ,AAA29771.1 ,BAA31183.1 ,AAA29779.1 ,CAA12110.1 ,AAA29770.1 ,BAA31177.1 ,BAA31167.1 ,AAA29767.1 ,AAA29775.1 ,AAB03226.1 ,CAA36267.1 ,P17201 ,CAB40275.1 ,BAA31170.1 ,CAA42177.1 ,AAB51545.1 ,AAF25352.1 ,AAF19243.1 ,CAA46541.1 ,CAA28951.1 ,AAA35619.1 ,AAA35618.1 ,AAC23952.1 ,AAD01978.1 ,BAA82222.1 ,AAA29776.1 ,AAF51552.1 ,AAB38802.1 ,AAA26794.1 ,AAA03201.1 ,CAA76129.1 ,AAF22492.2 ,AAA28032.2 ,AAB51551.1 ,CAA23544.1 ,CAA29981.1 ,AAB38801.1 ,AAD38797.1 ,AAB51549.1 ,AAB65643.1 ,CAB46203.1 ,AAB51555.1 ,AAC97484.1 ,AAB63303.1 ,AAB51549.1 ,AAB65643.1 ,CAB37806.1 ,AAD50541.1 ,AAF01565.1 ,AAB51555.1 ,AAC72024.1 ,BAA23627.1 ,CAA79152.1 ,AAD26128.1 ,AAB65643.1 ,AAC47463.1 ,AAB51559.1 ,AAD50543.1 ,AAB51560.1 ,Q04857 ,BAA07536.1 ,AAF69610.1 ,Q61704 ,AAB38774.1 ,AAB63302.1 ,AAB59564.1 ,AAC47462.1 ,BAA13940.1 ,CAA23883.1 ,CAA67576.1 ,CAA72309.1 ,Q63416 ,AAC23668.1 ,AAA93385.1 ,AAC47463.1 ,AAB51559.1 ,AAD50543.1 ,AAB51560.1 ,Q04857 ,BAA07536.1 ,AAF69610.1 ,Q61704 ,AAB38774.1 ,AAB63302.1 ,AAB59564.1 ,AAC47462.1 ,BAA13940.1 ,CAA23883.1 ,CAA67576.1 ,CAA72309.1 ,Q63416 ,AAC23668.1 ,AAA93385.1 ,

FIG. 3

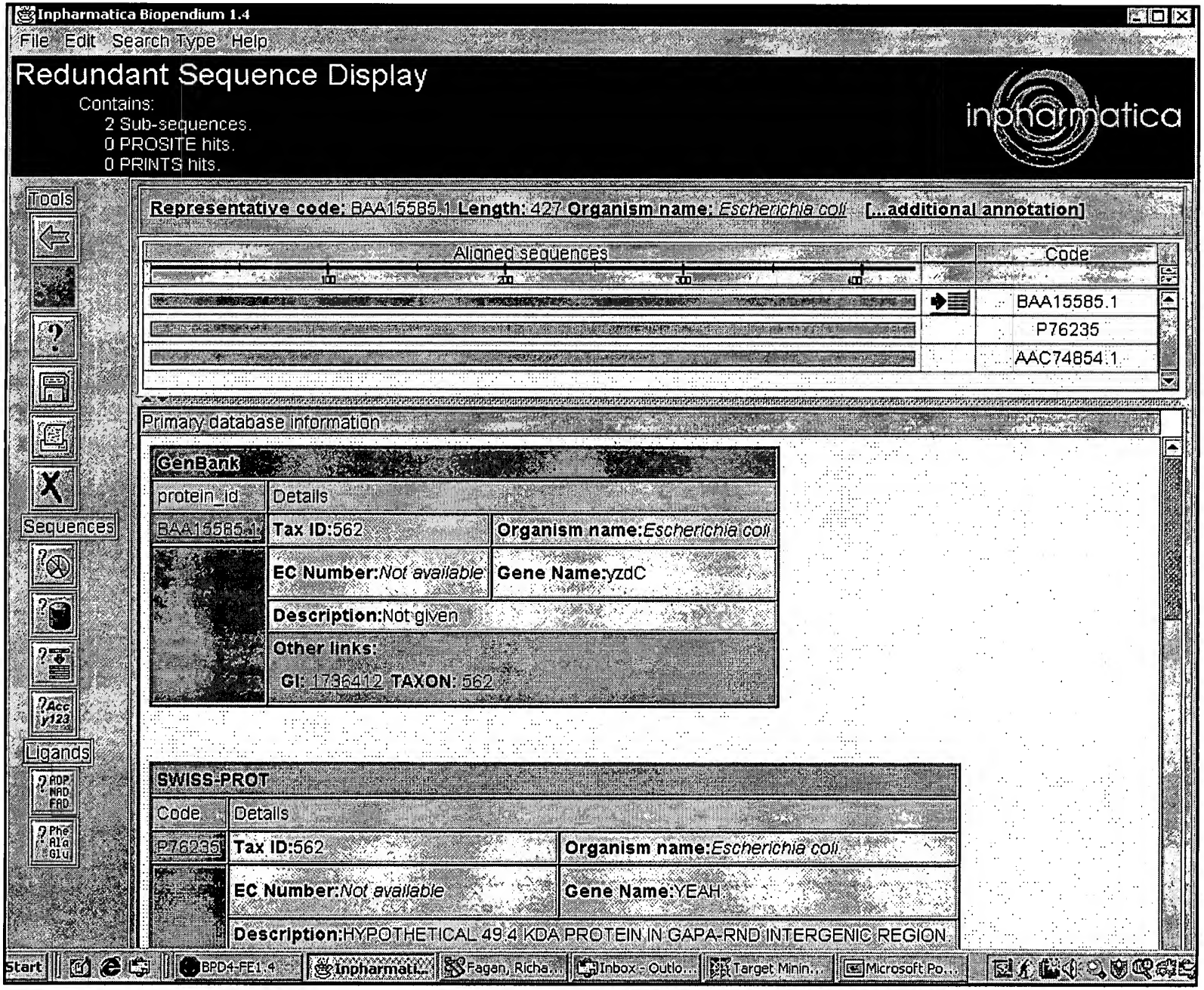

Inpharmatica Biopendium 1.4
File Edit Search Type Help
Redundant Sequence Display
Contains:
2 Sub-sequences.
0 PROSITE hits.
0 PRINTS hits.
inpharmatica
Tools
Sequences
Ligands
Representative code: BAA15585.1 Length: 427 Organism name: Escherichia coli [...additional annotation]
Aligned sequences
Code
BAA15585.1
P76235
AAC74854.1
Primary database information
GenBank
protein_id
Details
BAA15585.1
Tax ID:562
Organism name:Escherichia coli
EC Number:Not available
Gene Name:yzdC
Description:Not given
Other links:
GI: 1736412 TAXON: 562
SWISS-PROT
Code
Details
P76235
Tax ID:562
Organism name:Escherichia coli
EC Number:Not available
Gene Name:YEAH
Description:HYPOTHETICAL 49.4 KDA PROTEIN IN GAPA-RND INTERGENIC REGION
Start
BPD4-FE1.4
Inpharmati...
Fagen, Richa...
Inbox - Outlo...
Target Minin...
Microsoft Po...

FIG. 4

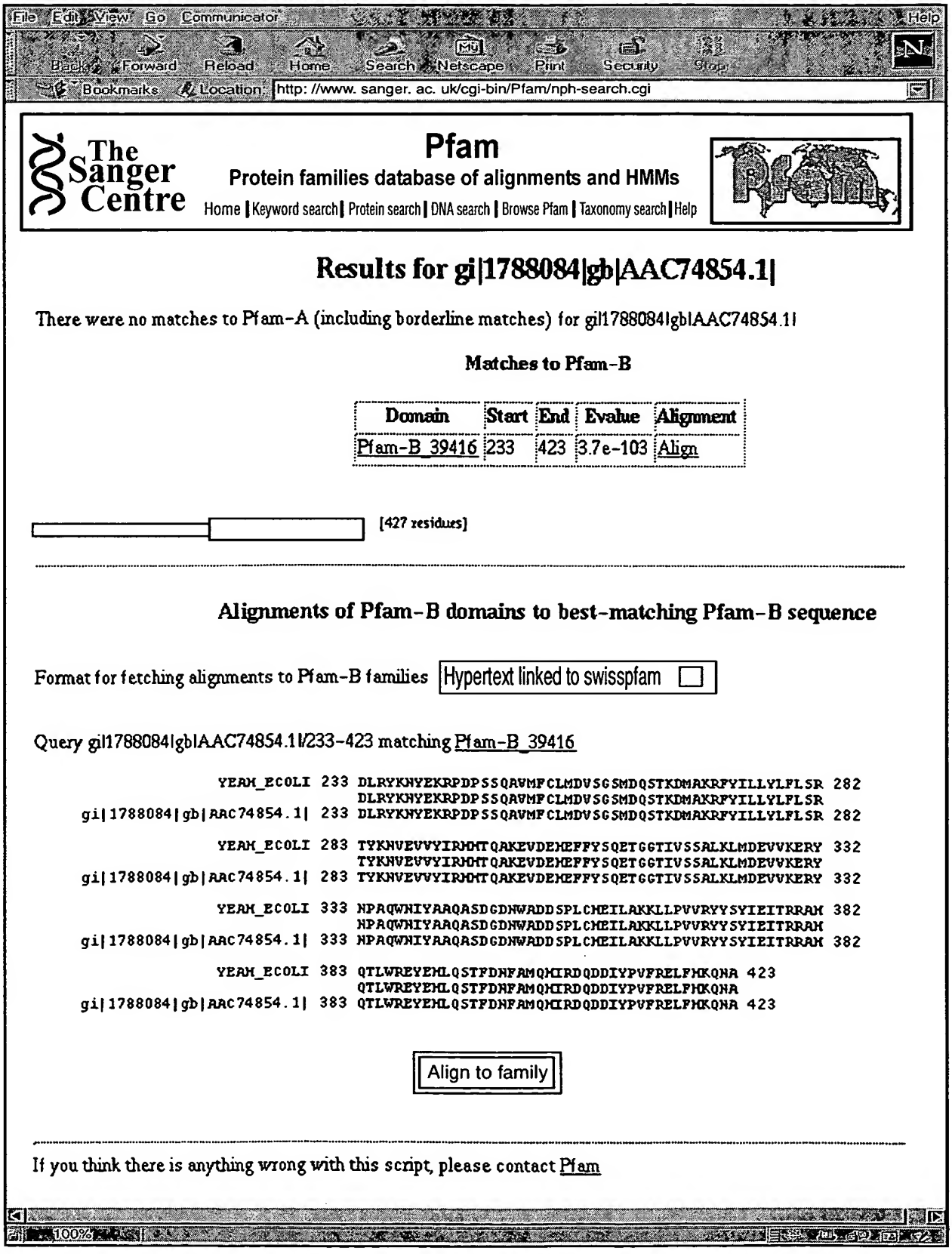
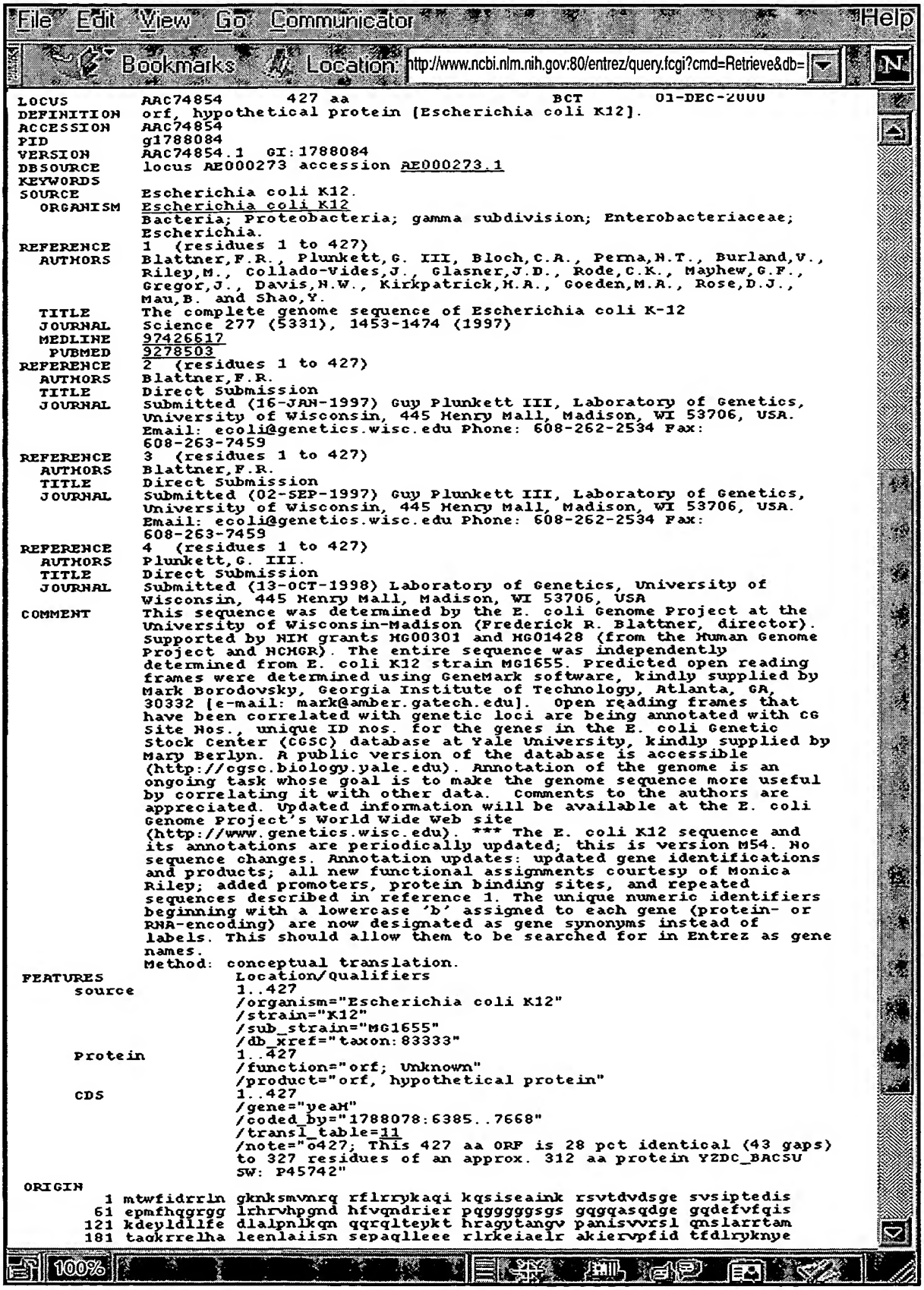

File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmarks Location: http: //www. sanger. ac. uk/cgi-bin/Pfam/nph-search.cgi
The Sanger Centre
Pfam
Protein families database of alignments and HMMs
Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help
Results for gi|1788084|gb|AAC74854.1|
There were no matches to Pfam-A (including borderline matches) for gi|1788084|gb|AAC74854.1|
Matches to Pfam-B
Domain Start End Evalue Alignment
Pfam-B_39416 233 423 3.7e-103 Align
[427 residues]
Alignments of Pfam-B domains to best-matching Pfam-B sequence
Format for fetching alignments to Pfam-B families Hypertext linked to swisspfam
Query gi|1788084|gb|AAC74854.1|/233-423 matching Pfam-B_39416
YEAH_ECOLI 233 DLRYKNYEKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRFYILLYLFLSR 282
DLRYKNYEKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRFYILLYLFLSR
gi|1788084|gb|AAC74854.1| 233 DLRYKNYEKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRFYILLYLFLSR 282
YEAH_ECOLI 283 TYKNVEVVYIRMMTQAKEVDEMEFFYSQETGGTIVSSALKLMDEVVKERY 332
TYKNVEVVYIRMMTQAKEVDEMEFFYSQETGGTIVSSALKLMDEVVKERY
gi|1788084|gb|AAC74854.1| 283 TYKNVEVVYIRMMTQAKEVDEMEFFYSQETGGTIVSSALKLMDEVVKERY 332
YEAH_ECOLI 333 NPAQWNIYAAQASDGDNWADDSPLCMEILAKKLLPVVRYYSYIEITRRAM 382
NPAQWNIYAAQASDGDNWADDSPLCMEILAKKLLPVVRYYSYIEITRRAM
gi|1788084|gb|AAC74854.1| 333 NPAQWNIYAAQASDGDNWADDSPLCMEILAKKLLPVVRYYSYIEITRRAM 382
YEAH_ECOLI 383 QTLWREYEMLQSTFDNFAMQMIRDQDDIYPVFRELFMKQNA 423
QTLWREYEMLQSTFDNFAMQMIRDQDDIYPVFRELFMKQNA
gi|1788084|gb|AAC74854.1| 383 QTLWREYEMLQSTFDNFAMQMIRDQDDIYPVFRELFMKQNA 423
Align to family
If you think there is anything wrong with this script, please contact Pfam
100%

FIG. 5





File Edit View Go Communicator Help

Bookmarks Location: http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=

```
LOCUS       AAC74854      427 aa                    BCT     01-DEC-2000
DEFINITION  orf, hypothetical protein [Escherichia coli K12].
ACCESSION   AAC74854
PID         g1788084
VERSION     AAC74854.1  GI:1788084
DBSOURCE    locus AE000273 accession AE000273.1
KEYWORDS
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (residues 1 to 427)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (residues 1 to 427)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (residues 1 to 427)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (residues 1 to 427)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu].  Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data.  Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..427
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     Protein         1..427
                     /function="orf; Unknown"
                     /product="orf, hypothetical protein"
     CDS             1..427
                     /gene="yeaM"
                     /coded_by="1788078:6385..7668"
                     /transl_table=11
                     /note="0427; This 427 aa ORF is 28 pct identical (43 gaps)
                     to 327 residues of an approx. 312 aa protein YZDC_BACSU
                     SW: P45742"
ORIGIN
        1 mtwfidrrln gknksmvnrq rflrrykaqi kqsiseaink rsvtdvdsge svsiptedis
       61 epmfhqgrgg lrhrvhpgnd hfvqndrier pqgggggsgs gqgqasqdge gqdefvfqis
      121 kdeyldllfe dlalpnlkqn qqrqlteykt hragytangv panisvvrsl qnslarrtam
      181 taakrrelha leenlaiisn sepaqlleee rlrkeiaelr akiervpfid tfdlryknye
```

100%

FIG. 6A

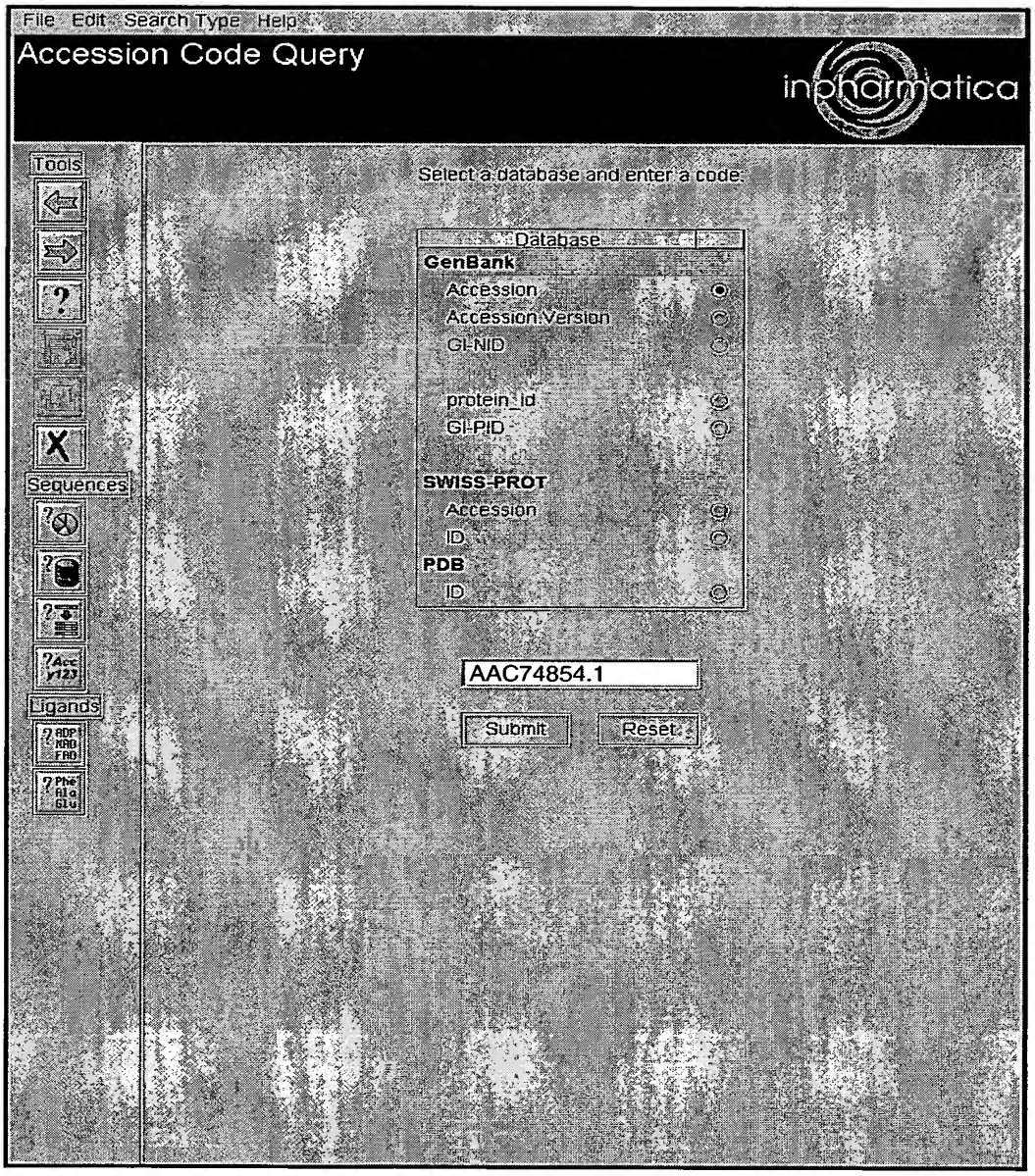

File Edit Search Type Help
Accession Code Query
inpharmatica
Tools
Sequences
Ligands
Select a database and enter a code.
Database
GenBank
Accession
Accession.Version
GI-NID
protein_id
GI-PID
SWISS-PROT
Accession
ID
PDB
ID
AAC74854.1
Submit
Reset

FIG. 6B

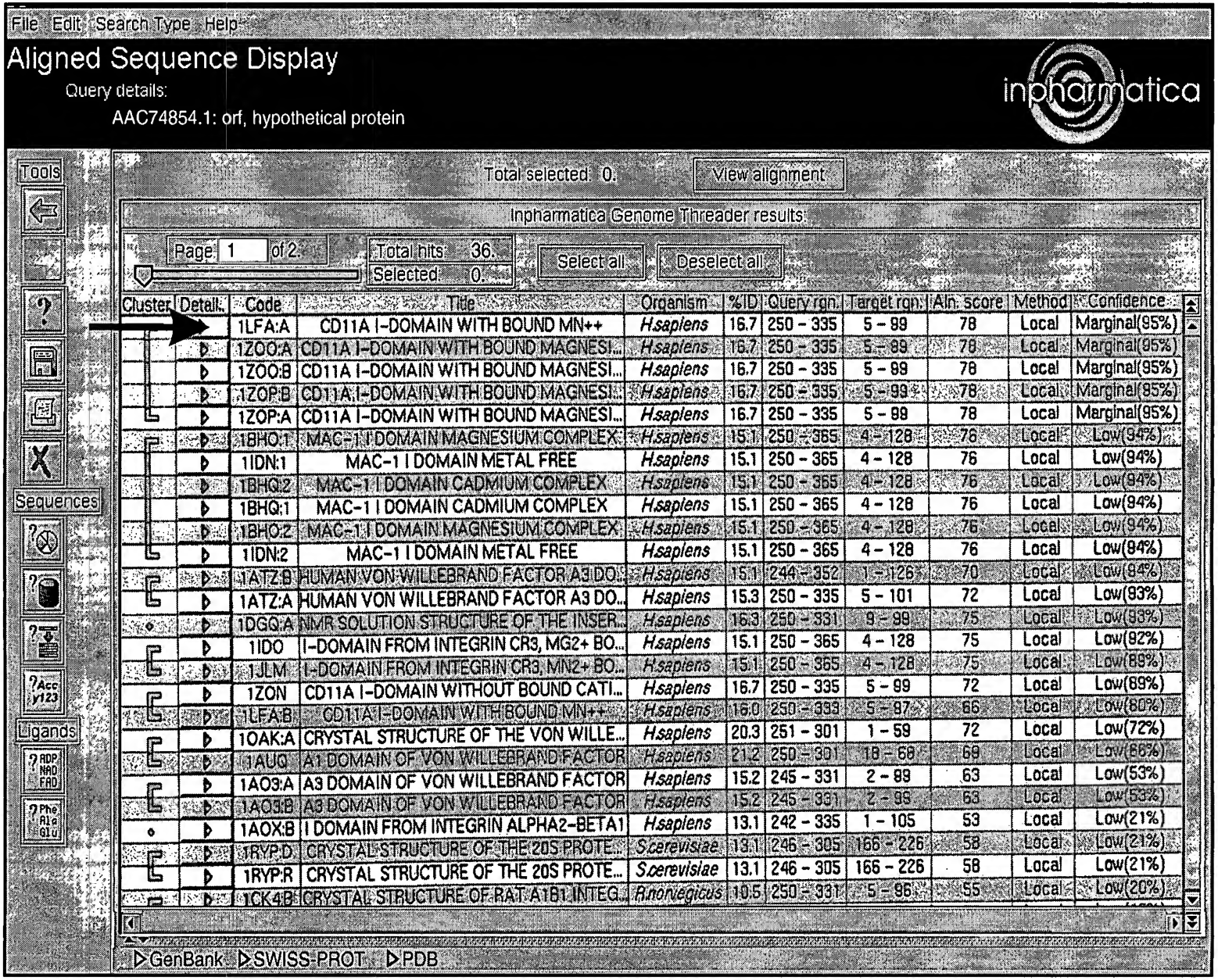
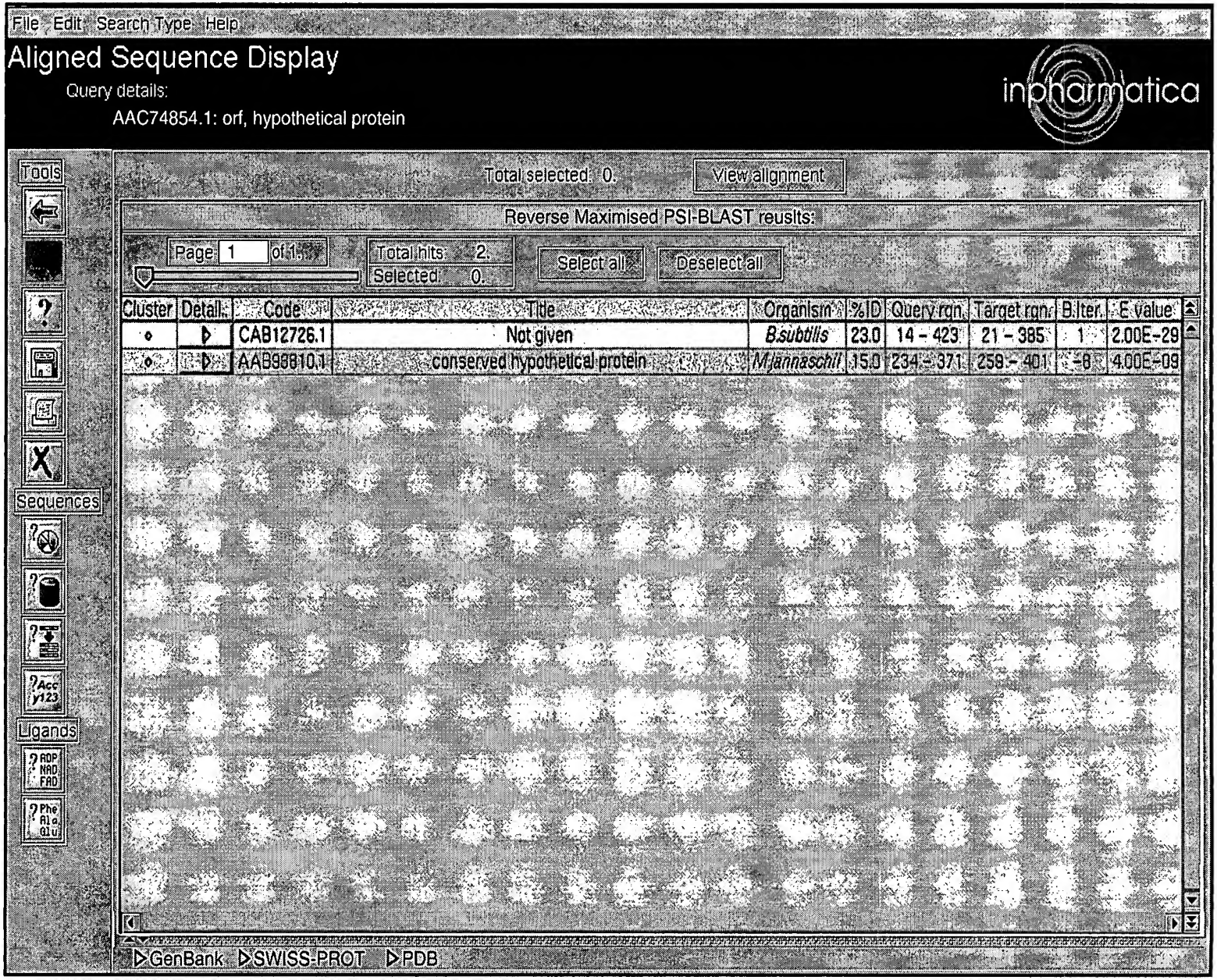

File Edit Search Type Help
Aligned Sequence Display
Query details:
AAC74854.1: orf, hypothetical protein
inpharmatica
Tools
Sequences
Ligands
Total selected: 0.
View alignment
Inpharmatica Genome Threader results
Page 1 of 2.
Total hits 36.
Selected 0.
Select all
Deselect all
Cluster Detail. Code Title Organism %ID Query rgn. Target rgn. Aln. score Method Confidence
1LFA:A CD11A I-DOMAIN WITH BOUND MN++ H.sapiens 16.7 250 - 335 5 - 99 78 Local Marginal(95%)
1ZOO:A CD11A I-DOMAIN WITH BOUND MAGNESI... H.sapiens 16.7 250 - 335 5 - 99 78 Local Marginal(95%)
1ZOO:B CD11A I-DOMAIN WITH BOUND MAGNESI... H.sapiens 16.7 250 - 335 5 - 99 78 Local Marginal(95%)
1ZOP:B CD11A I-DOMAIN WITH BOUND MAGNESI... H.sapiens 16.7 250 - 335 5 - 99 78 Local Marginal(95%)
1ZOP:A CD11A I-DOMAIN WITH BOUND MAGNESI... H.sapiens 16.7 250 - 335 5 - 99 78 Local Marginal(95%)
1BHO:1 MAC-1 I DOMAIN MAGNESIUM COMPLEX H.sapiens 15.1 250 - 365 4 - 128 76 Local Low(94%)
1IDN:1 MAC-1 I DOMAIN METAL FREE H.sapiens 15.1 250 - 365 4 - 128 76 Local Low(94%)
1BHQ:2 MAC-1 I DOMAIN CADMIUM COMPLEX H.sapiens 15.1 250 - 365 4 - 128 76 Local Low(94%)
1BHQ:1 MAC-1 I DOMAIN CADMIUM COMPLEX H.sapiens 15.1 250 - 365 4 - 128 76 Local Low(94%)
1BHO:2 MAC-1 I DOMAIN MAGNESIUM COMPLEX H.sapiens 15.1 250 - 365 4 - 128 76 Local Low(94%)
1IDN:2 MAC-1 I DOMAIN METAL FREE H.sapiens 15.1 250 - 365 4 - 128 76 Local Low(94%)
1ATZ:B HUMAN VON WILLEBRAND FACTOR A3 DO... H.sapiens 15.1 244 - 352 1 - 126 70 Local Low(84%)
1ATZ:A HUMAN VON WILLEBRAND FACTOR A3 DO... H.sapiens 15.3 250 - 335 5 - 101 72 Local Low(93%)
1DGQ:A NMR SOLUTION STRUCTURE OF THE INSER... H.sapiens 16.3 250 - 331 9 - 99 75 Local Low(93%)
1IDO I-DOMAIN FROM INTEGRIN CR3, MG2+ BO... H.sapiens 15.1 250 - 365 4 - 128 75 Local Low(92%)
1JLM I-DOMAIN FROM INTEGRIN CR3, MN2+ BO... H.sapiens 15.1 250 - 365 4 - 128 75 Local Low(89%)
1ZON CD11A I-DOMAIN WITHOUT BOUND CATI... H.sapiens 16.7 250 - 335 5 - 99 72 Local Low(89%)
1LFA:B CD11A I-DOMAIN WITH BOUND MN++ H.sapiens 16.0 250 - 333 5 - 97 66 Local Low(80%)
1OAK:A CRYSTAL STRUCTURE OF THE VON WILLE... H.sapiens 20.3 251 - 301 1 - 59 72 Local Low(72%)
1AUQ A1 DOMAIN OF VON WILLEBRAND FACTOR H.sapiens 21.2 250 - 301 18 - 69 69 Local Low(86%)
1AO3:A A3 DOMAIN OF VON WILLEBRAND FACTOR H.sapiens 15.2 245 - 331 2 - 89 63 Local Low(53%)
1AO3:B A3 DOMAIN OF VON WILLEBRAND FACTOR H.sapiens 15.2 245 - 331 2 - 96 63 Local Low(53%)
1AOX:B I DOMAIN FROM INTEGRIN ALPHA2-BETA1 H.sapiens 13.1 242 - 335 1 - 105 53 Local Low(21%)
1RYP:D CRYSTAL STRUCTURE OF THE 20S PROTE... S.cerevisiae 13.1 246 - 305 166 - 226 58 Local Low(21%)
1RYP:R CRYSTAL STRUCTURE OF THE 20S PROTE... S.cerevisiae 13.1 246 - 305 166 - 226 58 Local Low(21%)
1CK4:B CRYSTAL STRUCTURE OF RAT A1B1 INTEG... R.norvegicus 10.5 250 - 331 5 - 96 55 Local Low(20%)
GenBank SWISS-PROT PDB

FIG. 6C


File Edit Search Type Help
Aligned Sequence Display
Query details:
AAC74854.1: orf, hypothetical protein
inpharmatica
Tools
Total selected: 0.
View alignment
Reverse Maximised PSI-BLAST reuslts:
Page 1 of 1
Total hits: 2.
Selected: 0.
Select all
Deselect all
Cluster
Detail
Code
Title
Organism
%ID
Query rgn
Target rgn
B.Iter.
E value
CAB12726.1
Not given
B.subtilis
23.0
14 - 423
21 - 385
1
2.00E-29
AAB98810.1
conserved hypothetical protein
M.jannaschii
15.0
234 - 371
258 - 401
8
4.00E-09
Sequences
Ligands
GenBank
SWISS-PROT
PDB

# FIG. 7

AlEye output (January 4, 2002 3:07 PM)

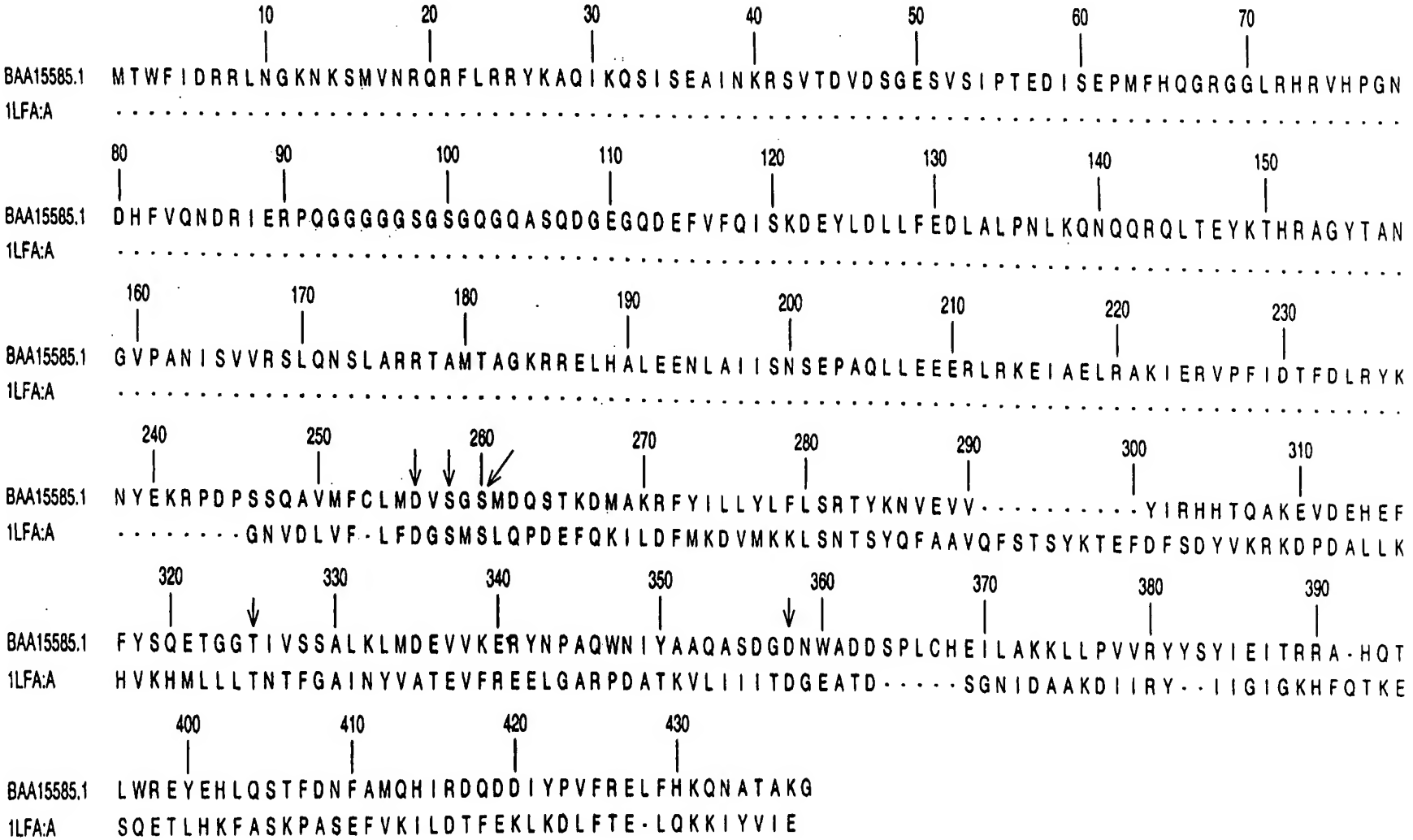
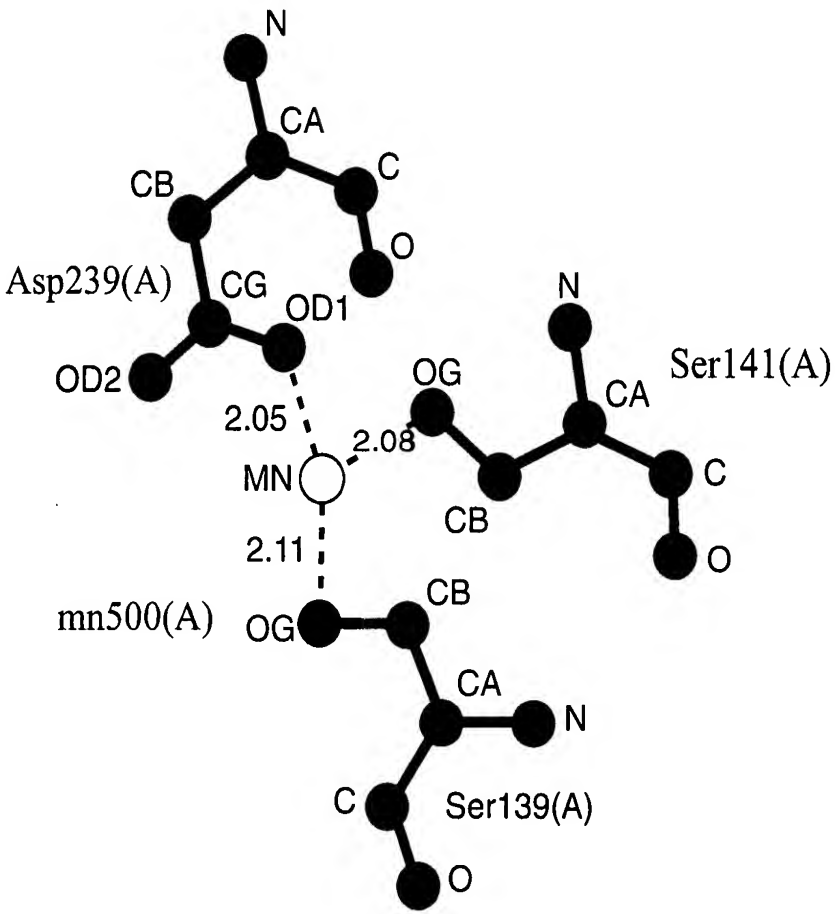


FIG. 8A




1lfa: MN500 Chain [A]

FIG. 8B

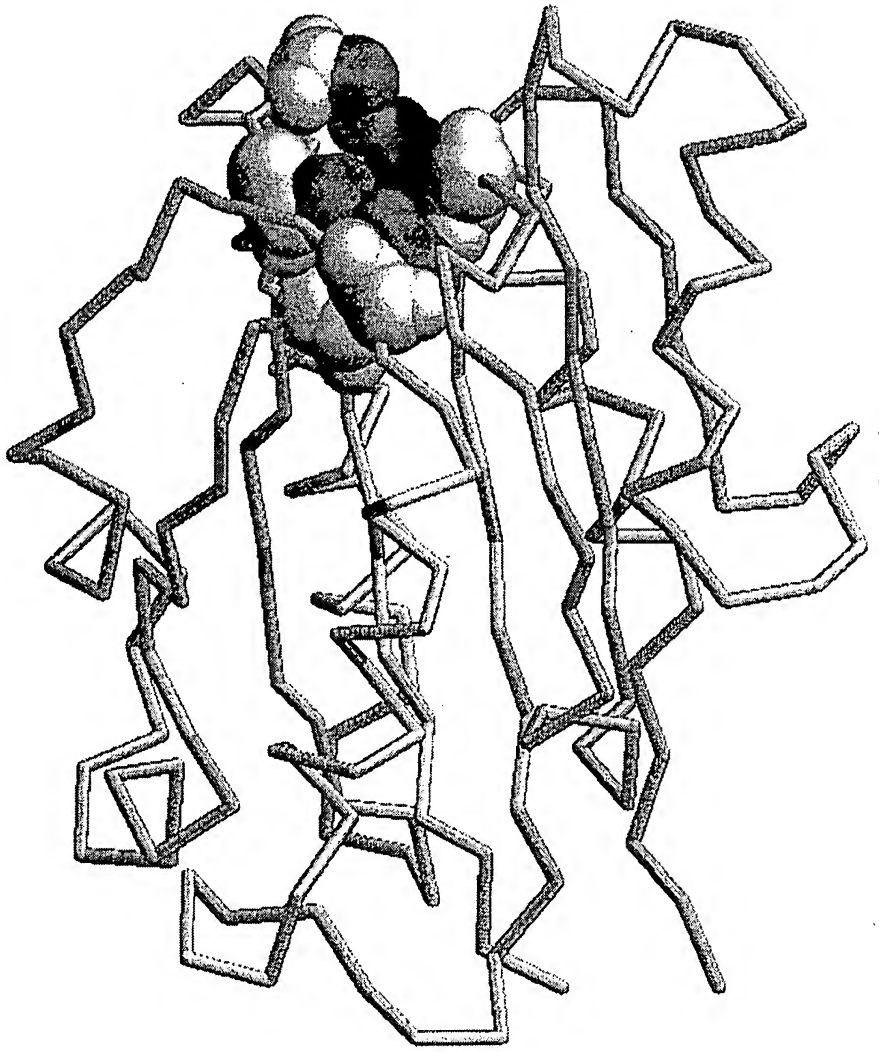
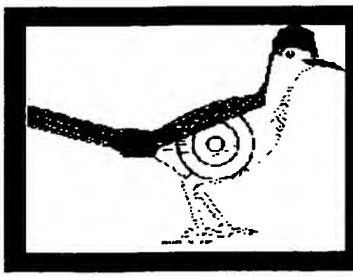


FIG. 9



File Edit View Go Communicator Help

Bookmarks Location: http: //victoria. inpharmatica.co.uk/~volker/BPD3target.html



Target Mining Interface

inpharmatica

## Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): 1AOX and chain (e.g. B): A

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

## Select Database

Release: DEVF9 = BPD3

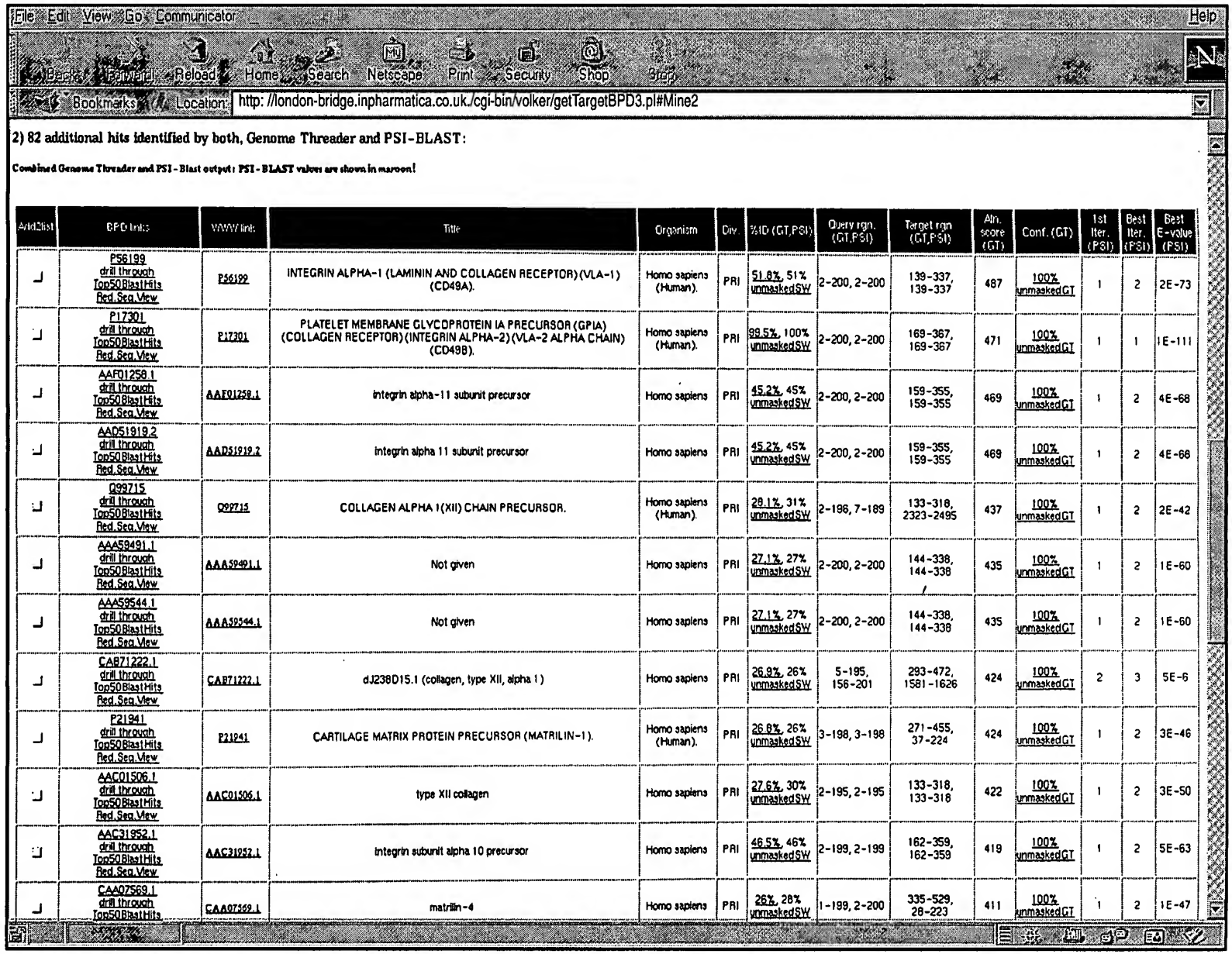
## Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

None

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

FIG. 10A




**2) 82 additional hits identified by both, Genome Threader and PSI-BLAST:**

**Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!**

| Add2list | BPD links | WWW link | Title | Organism | Div. | %ID (GT,PSI) | Query rgn. (GT,PSI) | Target rgn (GT,PSI) | Aln. score (GT) | Conf. (GT) | 1st Iter. (PSI) | Best Iter. (PSI) | Best E-value (PSI) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☐ | P56199 drill through Top50BlastHits Red.Seq.View | P56199 | INTEGRIN ALPHA-1 (LAMININ AND COLLAGEN RECEPTOR) (VLA-1) (CD49A). | Homo sapiens (Human). | PRI | 51.8%, 51% unmaskedSW | 2-200, 2-200 | 139-337, 139-337 | 487 | 100% unmaskedGT | 1 | 2 | 2E-73 |
| ☐ | P17301 drill through Top50BlastHits Red.Seq.View | P17301 | PLATELET MEMBRANE GLYCOPROTEIN IA PRECURSOR (GPIA) (COLLAGEN RECEPTOR) (INTEGRIN ALPHA-2) (VLA-2 ALPHA CHAIN) (CD49B). | Homo sapiens (Human). | PRI | 99.5%, 100% unmaskedSW | 2-200, 2-200 | 169-367, 169-367 | 471 | 100% unmaskedGT | 1 | 1 | 1E-111 |
| ☐ | AAF01258.1 drill through Top50BlastHits Red.Seq.View | AAF01258.1 | integrin alpha-11 subunit precursor | Homo sapiens | PRI | 45.2%, 45% unmaskedSW | 2-200, 2-200 | 159-355, 159-355 | 469 | 100% unmaskedGT | 1 | 2 | 4E-68 |
| ☐ | AAD51919.2 drill through Top50BlastHits Red.Seq.View | AAD51919.2 | integrin alpha 11 subunit precursor | Homo sapiens | PRI | 45.2%, 45% unmaskedSW | 2-200, 2-200 | 159-355, 159-355 | 469 | 100% unmaskedGT | 1 | 2 | 4E-68 |
| ☐ | Q99715 drill through Top50BlastHits Red.Seq.View | Q99715 | COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR. | Homo sapiens (Human). | PRI | 28.1%, 31% unmaskedSW | 2-196, 7-189 | 133-318, 2323-2495 | 437 | 100% unmaskedGT | 1 | 2 | 2E-42 |
| ☐ | AAA59491.1 drill through Top50BlastHits Red.Seq.View | AAA59491.1 | Not given | Homo sapiens | PRI | 27.1%, 27% unmaskedSW | 2-200, 2-200 | 144-338, 144-338 | 435 | 100% unmaskedGT | 1 | 2 | 1E-60 |
| ☐ | AAA59544.1 drill through Top50BlastHits Red.Seq.View | AAA59544.1 | Not given | Homo sapiens | PRI | 27.1%, 27% unmaskedSW | 2-200, 2-200 | 144-338, 144-338 | 435 | 100% unmaskedGT | 1 | 2 | 1E-60 |
| ☐ | CAB71222.1 drill through Top50BlastHits Red.Seq.View | CAB71222.1 | dJ238D15.1 (collagen, type XII, alpha 1) | Homo sapiens | PRI | 26.9%, 26% unmaskedSW | 5-195, 156-201 | 293-472, 1581-1626 | 424 | 100% unmaskedGT | 2 | 3 | 5E-6 |
| ☐ | P21941 drill through Top50BlastHits Red.Seq.View | P21941 | CARTILAGE MATRIX PROTEIN PRECURSOR (MATRILIN-1). | Homo sapiens (Human). | PRI | 26.6%, 26% unmaskedSW | 3-198, 3-198 | 271-455, 37-224 | 424 | 100% unmaskedGT | 1 | 2 | 3E-46 |
| ☐ | AAC01506.1 drill through Top50BlastHits Red.Seq.View | AAC01506.1 | type XII collagen | Homo sapiens | PRI | 27.6%, 30% unmaskedSW | 2-195, 2-195 | 133-318, 133-318 | 422 | 100% unmaskedGT | 1 | 2 | 3E-50 |
| ☐ | AAC31952.1 drill through Top50BlastHits Red.Seq.View | AAC31952.1 | integrin subunit alpha 10 precursor | Homo sapiens | PRI | 46.5%, 46% unmaskedSW | 2-199, 2-199 | 162-359, 162-359 | 419 | 100% unmaskedGT | 1 | 2 | 5E-63 |
| ☐ | CAA07569.1 drill through Top50BlastHits | CAA07569.1 | matrilin-4 | Homo sapiens | PRI | 26%, 28% unmaskedSW | 1-199, 2-200 | 335-529, 28-223 | 411 | 100% unmaskedGT | 1 | 2 | 1E-47 |

# FIG. 10B

| | | | | | | | | | | GT Confidence | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ⌟ | Top50BlastHits<br>Red.Seq.View | BAA91707.1 | Not given | sapiens | PRI | unmaskedSW | 80-195 | 36-141 | 142 | unmaskedGT | 3 | 5 | 2E-8 |
| ⌟ | AAA36795.1 drill through Top50BlastHits<br>Red.Seq.View | AAA36795.1 | undulin 2 | Homo sapiens | PRI | 34.7%, 21% unmaskedSW | 2-46, 139-196 | 3-51, 37-93 | 121 | 100% unmaskedGT | 2 | 2 | 8E-6 |
| ⌟ | AAD40367.1 drill through Top50BlastHits<br>Red.Seq.View | AAD40367.1 | calcium-activated chloride channel-2 | Homo sapiens | PRI | 15%, 15% unmaskedSW | 8-117, 8-117 | 312-421, 312-421 | 117 | 100% unmaskedGT | 5 | 5 | 1E-4 |
| ⌟ | CAA67559.1 drill through Top50BlastHits<br>Red.Seq.View | CAA67559.1 | collagen VI-alpha-1 chain | Homo sapiens | PRI | 21.1%, 21% unmaskedSW | 108-197, 108-197 | 7-92, 7-92 | 114 | 100% unmaskedGT | 3 | 3 | 5E-7 |
| → | AAC76768.1 drill through Top50BlastHits<br>Red.Seq.View | AAC76768.1 | orf, hypothetical protein | Escherichia coli | BCT | 13.2%, 11% unmaskedSW | 9-134, 7-186 | 267-384, 265-424 | 107 | 100% unmaskedGT | -4 | -5 | 3E-24 |
| ⌟ | CAB43000.1 drill through Top50BlastHits<br>Red.Seq.View | CAB43000.1 | tellurite resistance | Escherichia coli | BCT | 12.9%, 14% unmaskedSW | 8-192, 2-181 | 214-398, 208-380 | 102 | 100% unmaskedGT | -6 | -7 | 6E-22 |
| ⌟ | AAA60114.1 drill through Top50BlastHits<br>Red.Seq.View | AAA60114.1 | platelet membrane glycoprotein IIb | Homo sapiens | PRI | 14%, 13% unmaskedSW | 157-199, 157-199 | 145-187, 145-187 | 111 | 98.52% unmaskedGT<br>reverse Hit | -7 | -8 | 8E-5 |

608 out of these 632 PSI-BLAST matches were identified using 'positive iterations':

FIG. 10C

P17301,1AOX:B,AAB59255.1,CAA82877.1,CAA78125.1,P56199,CAA36384.1,BAA23160.1,1CK4:B,1CK4:A,1QC5:B,1QC5:A,AAF01258.1,AAD51919.2,AAC31952.1,CAA30479.1,AAA59544.1,AAB24821.1,1BHO:A,1BHO:B,1IDN:A,1BHQ:A,1BHQ:B,1IDN:B,AAB16869.1,1ILN,CAA50063.1,P32018,AAA99719.1,BAA00701.1,Q99715,1IDO,AAB38702.1,CAA06890.1,CAA06889.1,AAF21241.1,CAA07569.1,AAB38547.1,CAB46380.1,P51942,P20701,P05099,AAC31672.1,1DGO:A,Q91145,CAB71222.1,1LFA:A,1ZOO:B,1ZOO:A,1ZOP:B,1ZOP:A,Q99715,BAA00701.1,1LFA:B,AAA99719.1,AAD25884.1,AAA33426.1,AAD25885.1,Q99715,1ZON,AAC23663.1,CAB70853.1,AAC01506.1,CAA50063.1,P32018,AAA99719.1,AAC53163.1,CAA72402.1,AAF43107.1,P38570,AAD04919.1,AAB93766.1,BAA96034.1,AAC52142.1,BAA00701.1,CAA27972.1,AAB59512.1,AAA82929.1,AAC53163.1,CAB37653.1,CAA07569.1,CAA12110.1,CAB86185.1,CAA68283.1,CAB46380.1,CAB37847.1,AAB65641.1,AAC06229.1,AAF25353.1,AAB51548.1,AAD50539.1,CAB37672.1,CAB37665.1,CAA03885.1,CAB37852.1,P05099,AAB38702.1,AAB51556.1,AAB51557.1,P51942,CAB86200.1,CAB86203.1,CAB40275.1,CAB37666.1,CAB37856.1,AAB51554.1,CAB37651.1,AAB51553.1,AAB38798.1,AAA51620.1,AAB51550.1,CAA06889.1,AAC72024.1,CAB86202.1,CAA36267.1,AAF25351.1,AAD50540.1,CAB86227.1,AAD50544.1,AAD01701.1,AAD26126.1,AAB51546.1,AAD26124.1,AAB21220.2,AAF42927.1,CAB86184.1,AAB65642.1,CAB86188.1,AAC32844.1,CAB37670.1,CAB37668.1,AAD50542.1,AAB38800.1,AAD26125.1,AAB03226.1,AAC32845.1,AAF25354.1,CAB65984.1,CAB65984.1,CAB37843.1,AAB38802.1,AAB51552.1,CAB37667.1,AAB51549.1,AAC32846.1,Q99715,AAB51555.1,AAD29428.1,AAB38799.1,AAF25352.1,CAB37857.1,AAB51551.1,AAA36794.1,AAA28092.2,CAB37849.1,AAB65643.1,AAB38797.1,AAF23492.1,AAD01978.1,AAA03201.1,AAB51558.1,CAB37846.1,AAB51545.1,AAF42926.1,AAD50543.1,AAA03201.1,AAA03201.1,AAA99719.1,CAA36267.1,AAF42925.1,AAA03201.1,CAA71532.1,AAA96311.1,AAD01978.1,AAC23667.1,AAC62253.1,CAA27972.1,AAB59512.1,1ATE:B,1AO3:B,1AO3:A,CAA36267.1,AAA03201.1,AAD26127.1,AAD01978.1,CAA36267.1,AAD01701.1,AAD04919.1,1ATE:A,AAB93766.1,AAD01978.1,AAD40483.1,CAA36267.1,CAA97806.1,CAA36267.1,AAF01565.1,CAA36267.1,AAD01978.1,AAC23667.1,AAB63303.1,AAC62253.1,Q43405,AAA29076.1,AAC39949.1,AAA03201.1,AAF19243.1,AAC23668.1,AAC06229.1,AAD01978.1,BAA23627.1,AAA03201.1,AAA03201.1,AAC23667.1,CAA36267.1,AAD03350.1,AAC23667.1,AAC79682.1,AAD01978.1,CAA36267.1,AAC68807.1,BAA82322.1,CAA47744.1,CAA98248.1,BAA00701.1,AAC47957.1,AAA50715.1,AAC19251.1,AAB52325.1,AAF19243.1,BAA82322.1,AAC02556.1,AAB66593.1,AAD01978.1,AAC23667.1,AAA58965.1,Q02388,CAB37672.1,AAD01978.1,AAC23667.1,CAB37849.1,AAB51551.1,AAF25353.1,AAB38799.1,CAB37670.1,AAB38797.1,CAB86200.1,CAB58678.1,1AUQ,1OAK:A,AAB51558.1,CAB37653.1,AAB51557.1,AAB59512.1,CAA27972.1,AAD50539.1,CAB86202.1,AAD50540.1,AAD50541.1,CAB37667.1,AAB65641.1,CAB37665.1,AAB65642.1,AAB51549.1,CAB86188.1,AAB51547.1,CAB37843.1,AAF42927.1,AAF25354.1,AAF42925.1,AAD26125.1,CAB86184.1,AAB51555.1,AAD26126.1,AAD26127.1,AAB51550.1,CAB86185.1,CAB37857.1,AAB51552.1,CAB86203.1,AAC06229.1,CAB37668.1,BAA31172.1,CAB37846.1,AAA29773.1,CAA31440.1,AAD04919.1,CAB37852.1,AAA96311.1,CAB86227.1,AAB93766.1,AAB51559.1,AAA29774.1,AAA29776.1,AAA29770.1,AAA29772.1,CAB37847.1,AAD50543.1,AAB51560.1,AAB38798.1,AAC97484.1,BAA31171.1,BAA31173.1,BAA31174.1,BAA31182.1,BAA31190.1,BAA31188.1,BAA31178.1,AAA29778.1,BAA31167.1,BAA31193.1,BAA31177.1,BAA31176.1,BAA31195.1,BAA31194.1,BAA31169.1,BAA31175.1,AAC18657.1,BAA31183.1,AAA29771.1,BAA31187.1,AAB38802.1,AAC47463.1,AAA29767.1,AAA29775.1,BAA31170.1,AAB21220.2,AAB51549.1,AAA03201.1,CAB37856.1,Q43405,AAC39949.1,AAB51554.1,AAC46961.1,AAB38801.1,CAB38978.1,AAC47462.1,CAA33891.1,AAA35618.1,AAA35619.1,AAD50544.1,CAA46541.1,CAA73140.1,AAC32844.1,AAF25352.1,AAB34889.1,AAC23667.1,AAF25351.1,AAC32845.1,CAB37651.1,AAD26124.1,AAB63302.1,AAB51546.1,AAC84162.1,CAB37666.1,AAB51556.1,CAA79152.1,CAA33144.1,CAA39981.1,AAA82929.1,AAA37380.1,AAA37381.1,AAB51545.1,AAB97607.1,BAA34707.1,Q04857,BAA07536.1,AAB51547.1,AAB51553.1,AAF69610.1,AAD50542.1,CAA33889.1,CAA67576.1,BAA02763.1,CAB03057.1,AAF42926.1,CAA51389.1,AAC32846.1,BAA06179.1,AAC47957.1,CAA33888.1,Q04857,BAA82322.1,AAB59954.1,AAC47461.1,CAB03055.1,CAB03054.1,AAB65643.1,BAA88322.1,CAB03056.1,CAB03059.1,CAB38978.1,AAC46961.1,BAA82322.1,CAB71222.1,AAC84160.1,AAB59954.1,CAB03058.1,CAA79153.1,AAB24637.1,AAB38800.1,AAB59954.1,CAB46421.1,CAA46541.1,BAA02853.1,CAA33891.1,CAA67576.1,CAA79153.1,Q04857,CAA88985.1,CAA79152.1,AAC46572.1,AAF00021.1,CAA33144.1,CAA39981.1,AAC24429.1,CAB75310.1,AAC23667.1,CAB38978.1,AAC46961.1,CAB02952.1,AAB51560.1,CAA33890.1,AAA58965.1,Q02388,CAA36267.1,AAA03201.1,AAF70042.1,CAB03680.1,CAA33889.1,AAD30859.1,CAA33144.1,CAA79156.1,AAA03201.1,CAB38978.1,AAC65240.1,BAA23627.1,CAA46541.1,P34393,AAD50541.1,CAA91447.1,AAD01978.1,BAA82322.1,AAC17639.1,AAC23667.1,CAA79156.1,AAD16289.1,AAB04590.2,CAA61109.1,AAB38801.1,CAB38978.1,AAC65016.1,AAB63363.1,AAA36795.1,AAC46961.1,AAC46961.1,AAC46572.1,AAF12450.1,AAD50541.1,AAA81430.1,AAA50715.1,CAB38978.1,AAC46961.1,AAC19251.1,AAA36795.1,CAA36267.1,Q91145,Q61703,BAA13939.1,P19823,CAA72308.1,CAA30160.1,Q61704,CAA72155.1,BAA78416.1,BAA13940.1,AAC79982.1,Q63416,AAC35003.1,AAC36289.1,P54290,AAB50142.1,AAB50141.1,AAB50140.1,AAB50139.1,AAB50138.1,CAA49279.1,CAA45188.1,AAA81562.1,AAF03259.1,BAA34706.1,CAA72309.1,BAA13938.1,AAC48511.1,AAB46821.1,P54289,AAF12731.1,BAA07632.1,CAA49841.1,AAB86529.1,AAB86531.1,AAB86530.1,CAB86192.1,AAC70914.1,CAB86193.1,CAA88984.1,AAD48398.1,Q55874,AAC25786.1,CAB77598.1,P76481,AAF53476.1,AAF44957.1,BAA12207.1,AAF13350.1,CAB10840.1,CAA88986.1,AAA82911.1,AAC79610.1,CAA88991.1,AAB63940.1,BAA08371.1,CAA88988.1,CAA91030.1,AAA36225.2,CAB08456.1,AAD30858.1,AAB19093.1,AAC72024.1,CAA16338.1,CAA34611.1,AAC05096.1,CAB59466.1,BAA80016.1,AAC79447.1,AAF39956.1,AAF10487.1,AAC65716.1,CAB11610.1,CAA10812.1,BAA91707.1,AAD34763.1,CAB03681.1,CAA67559.1,AAD03492.1,CAB38978.1,AAC46961.1,AAB51559.1,AAC95429.1,CAA09423.1,BAA33743.1,CAB75878.1,CAB75962.1,AAF58335.1,CAA97671.1,AAF53505.1,AAF44988.1,CAB71206.1,AAF27140.1,CAA47439.1,AAC83699.1,CAB76889.1,BAA25490.1,BAA20346.1,AAF22895.1,CAB69268.1,AAD52031.1,BAA16787.1,CAB81791.1,O22437,AAC83698.1,BAA18152.1,AAF69021.1,AAC75134.1,AAA60558.1,AAB85061.1,O50313,AAC75182.1,CAA16691.1,CAB80454.1,AAD41583.1,AAF58612.1,AAA27969.1,BAA76532.1,AAD31445.2,AAC44737.1,CAA77537.1,O34845,AAC84032.1,AAC62491.1,AAD15576.1,AAD47250.1,CAB13830.1,CAA30091.1,AAB98810.1,BAA79575.1,Q60384,BAA61542.1,AAF11936.1,AAD40367.1,AAB89815.1,AAA68973.1,BAA10368.1,AAF23117.1,AAF27914.1,CAB52078.1,BAA20761.1,AAD42192.2,AAB84957.1,

24 out of these 632 PSI-BLAST matches were identified using 'negative iterations':

BAA82333.1,CAA12066.1,CAB50193.1,CAB43000.1,CAB01991.1,CAA17374.1,BAA07331.1,AAC76768.1,AAB86281.1,CAB76247.1,AAC45374.1,AAB80774.1,AAA86852.2,AAB98912.1,AAA93386.1,AAA93386.1,AAC16991.1,AAF57485.1,BAA86551.1,AAB41649.1,AAA60114.1,AAC51323.1,AAC52946.1,Q01443,

FIG. 11

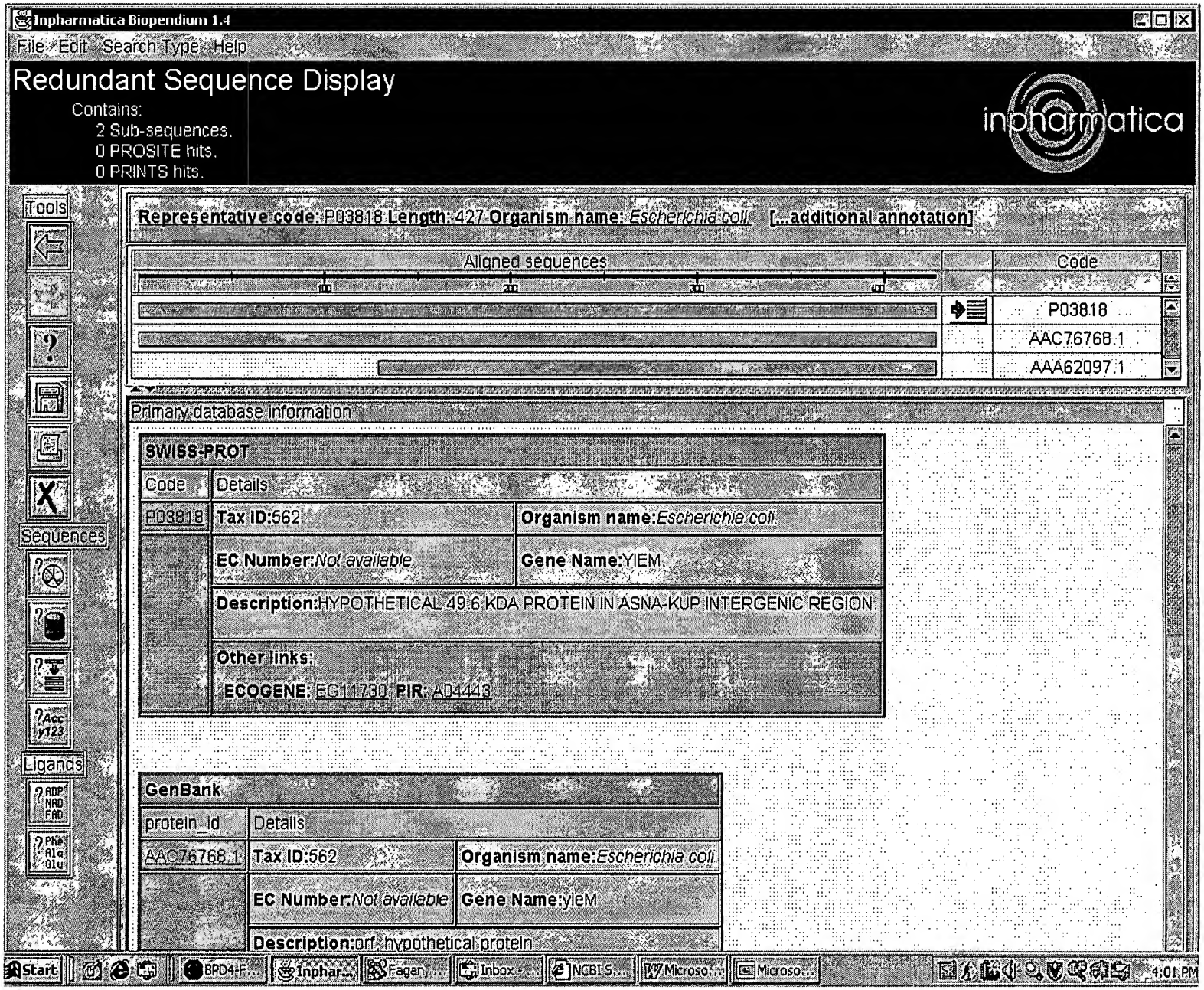

Inpharmatica Biopendium 1.4
File Edit Search Type Help
Redundant Sequence Display
Contains:
2 Sub-sequences.
0 PROSITE hits.
0 PRINTS hits.
inpharmatica
Tools
Sequences
Ligands
Representative code: P03818 Length: 427 Organism name: Escherichia coli. [...additional annotation]
Aligned sequences
Code
P03818
AAC76768.1
AAA62097.1
Primary database information
SWISS-PROT
Code
Details
P03818
Tax ID:562
Organism name:Escherichia coli.
EC Number:Not available
Gene Name:YIEM.
Description:HYPOTHETICAL 49.6 KDA PROTEIN IN ASNA-KUP INTERGENIC REGION.
Other links:
ECOGENE: EG11730 PIR: A04443
GenBank
protein_id
Details
AAC76768.1
Tax ID:562
Organism name:Escherichia coli
EC Number:Not available
Gene Name:yieM
Description:orf, hypothetical protein
Start
4:01 PM

FIG. 12

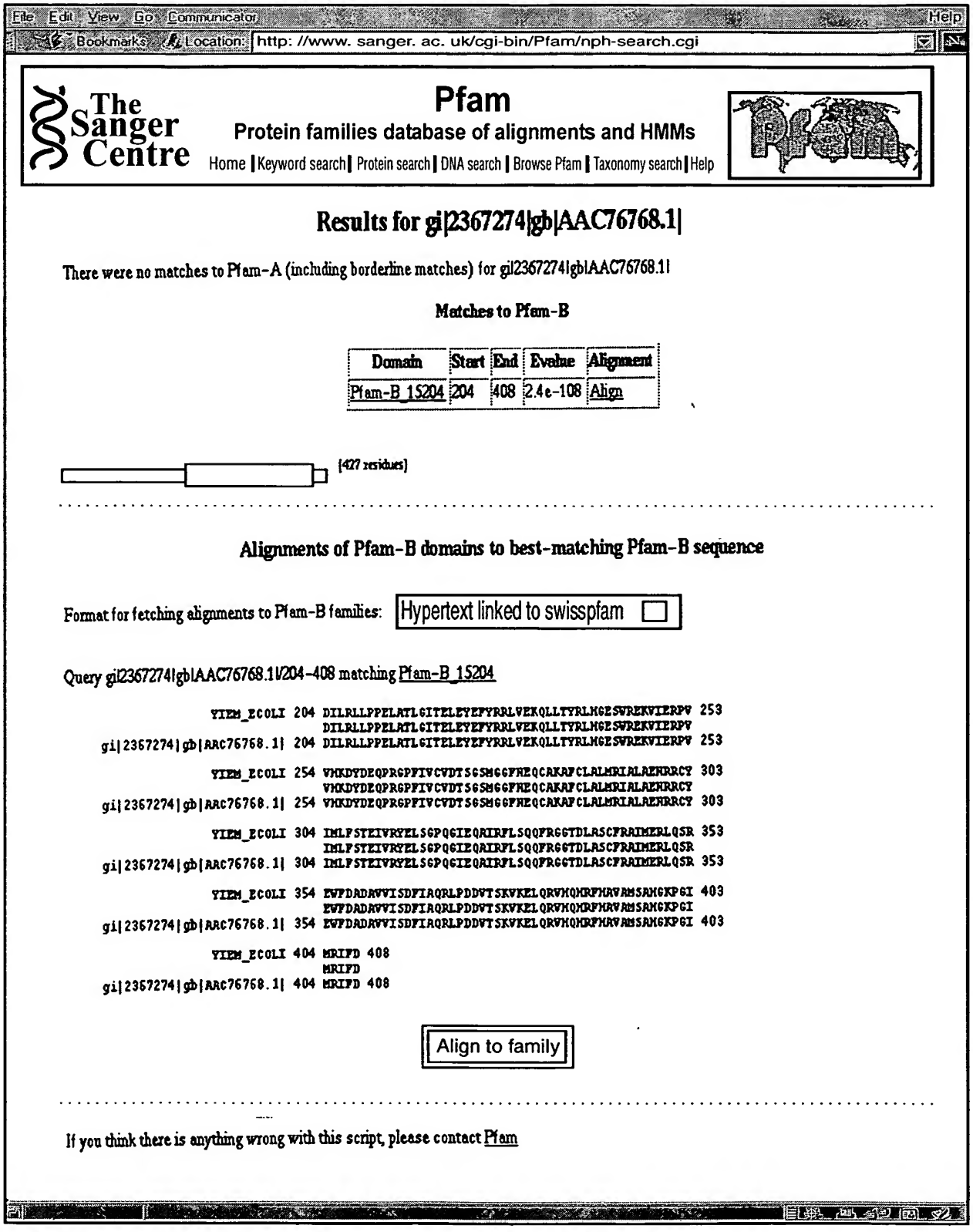

File Edit View Go Communicator Help
Bookmarks Location: http: //www. sanger. ac. uk/cgi-bin/Pfam/nph-search.cgi
The Sanger Centre
Pfam
Protein families database of alignments and HMMs
Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help
Results for gi|2367274|gb|AAC76768.1|
There were no matches to Pfam-A (including borderline matches) for gi|2367274|gb|AAC76768.1|
Matches to Pfam-B
Domain Start End Evalue Alignment
Pfam-B_15204 204 408 2.4e-108 Align
[427 residues]
Alignments of Pfam-B domains to best-matching Pfam-B sequence
Format for fetching alignments to Pfam-B families: Hypertext linked to swisspfam
Query gi|2367274|gb|AAC76768.1|/204-408 matching Pfam-B_15204
YIEM_ECOLI 204 DILRLLPPELATLGITELEYEFYRRLVEKQLLTYRLHGESWREKVIERPV 253
DILRLLPPELATLGITELEYEFYRRLVEKQLLTYRLHGESWREKVIERPV
gi|2367274|gb|AAC76768.1| 204 DILRLLPPELATLGITELEYEFYRRLVEKQLLTYRLHGESWREKVIERPV 253
YIEM_ECOLI 254 VMKDYDEQPRGPFIVCVDTSGSMGGFNEQCAKAFCLALMRIALAENRRCY 303
VMKDYDEQPRGPFIVCVDTSGSMGGFNEQCAKAFCLALMRIALAENRRCY
gi|2367274|gb|AAC76768.1| 254 VMKDYDEQPRGPFIVCVDTSGSMGGFNEQCAKAFCLALMRIALAENRRCY 303
YIEM_ECOLI 304 IMLPSTEIVRYELSGPQGIEQAIRFLSQQFRGGTDLASCFRAIMERLQSR 353
IMLPSTEIVRYELSGPQGIEQAIRFLSQQFRGGTDLASCFRAIMERLQSR
gi|2367274|gb|AAC76768.1| 304 IMLPSTEIVRYELSGPQGIEQAIRFLSQQFRGGTDLASCFRAIMERLQSR 353
YIEM_ECOLI 354 EWFDADAVVISDFIAQRLPDDVTSKVKELQRVHQHRFHAVAMSAHGKPGI 403
EWFDADAVVISDFIAQRLPDDVTSKVKELQRVHQHRFHAVAMSAHGKPGI
gi|2367274|gb|AAC76768.1| 354 EWFDADAVVISDFIAQRLPDDVTSKVKELQRVHQHRFHAVAMSAHGKPGI 403
YIEM_ECOLI 404 MRIFD 408
MRIFD
gi|2367274|gb|AAC76768.1| 404 MRIFD 408
Align to family
If you think there is anything wrong with this script, please contact Pfam

FIG. 13



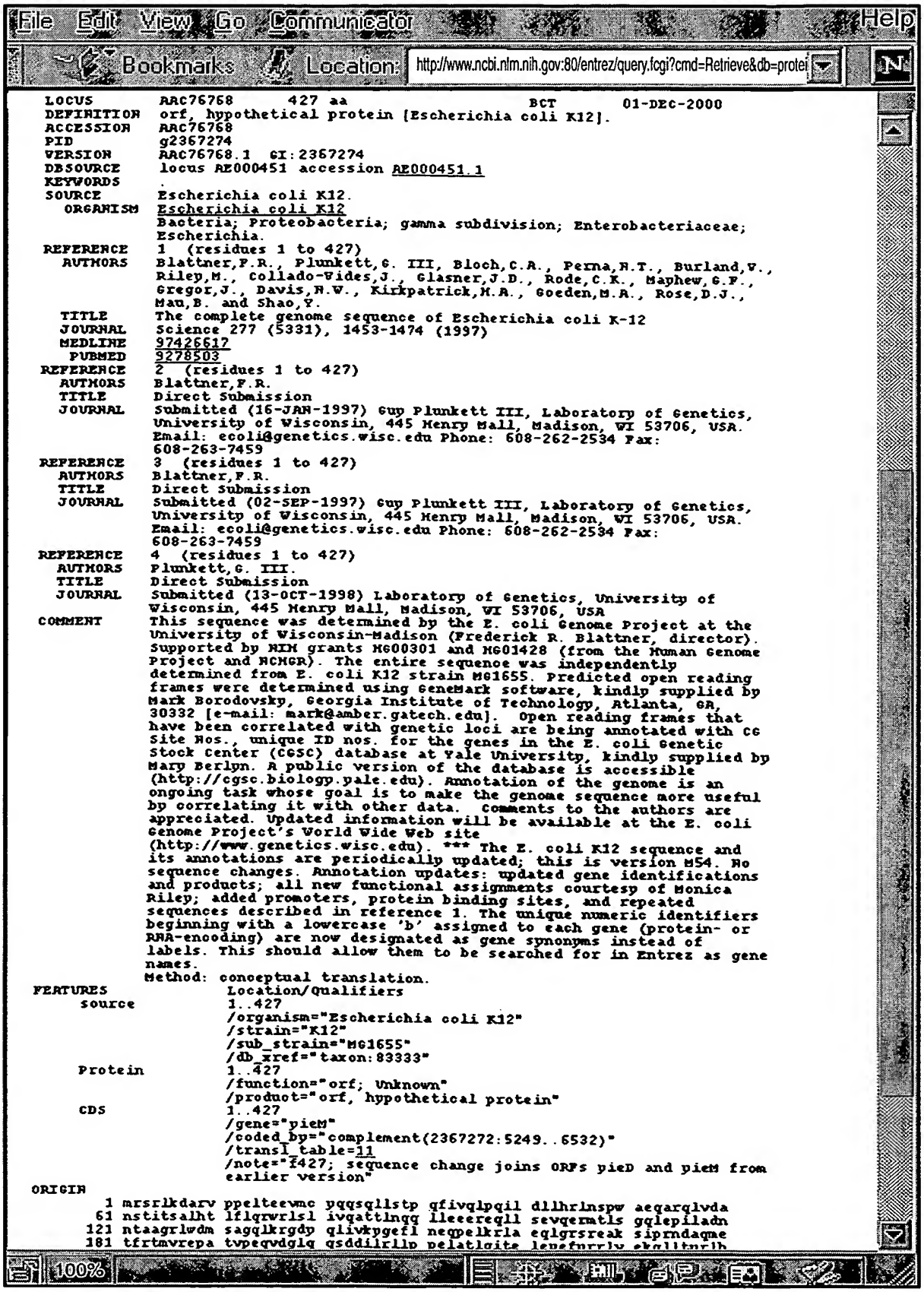

File Edit View Go Communicator Help
Bookmarks Location: http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=protei
LOCUS AAC76768 427 aa BCT 01-DEC-2000
DEFINITION orf, hypothetical protein [Escherichia coli K12].
ACCESSION AAC76768
PID g2367274
VERSION AAC76768.1 GI:2367274
DBSOURCE locus AE000451 accession AE000451.1
KEYWORDS .
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (residues 1 to 427)
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (residues 1 to 427)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 3 (residues 1 to 427)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 4 (residues 1 to 427)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and HCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
Method: conceptual translation.
FEATURES Location/Qualifiers
source 1..427
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
Protein 1..427
/function="orf; Unknown"
/product="orf, hypothetical protein"
CDS 1..427
/gene="yieM"
/coded_by="complement(2367272:5249..6532)"
/transl_table=11
/note="f427; sequence change joins ORFs yieD and yieM from earlier version"
ORIGIN
1 mrsrlkdarv ppelteevmc pqqsqllstp qfivqlpqil dllhrlnspw aeqarqlvda
61 nstitsalht lflqrwrlsl ivqattlnqq lleeereqll sevqermtls gqlepiladn
121 ntaagrlwdm sagqlkrgdp qlivkpgefl neqpelkrla eqlgrsreak siprndaqme
181 tfrtmvrepa tvpeqvdglq qsddilrllp pelatlgite lepefnrrlv ekqlltnrlh
100%

FIG. 14A

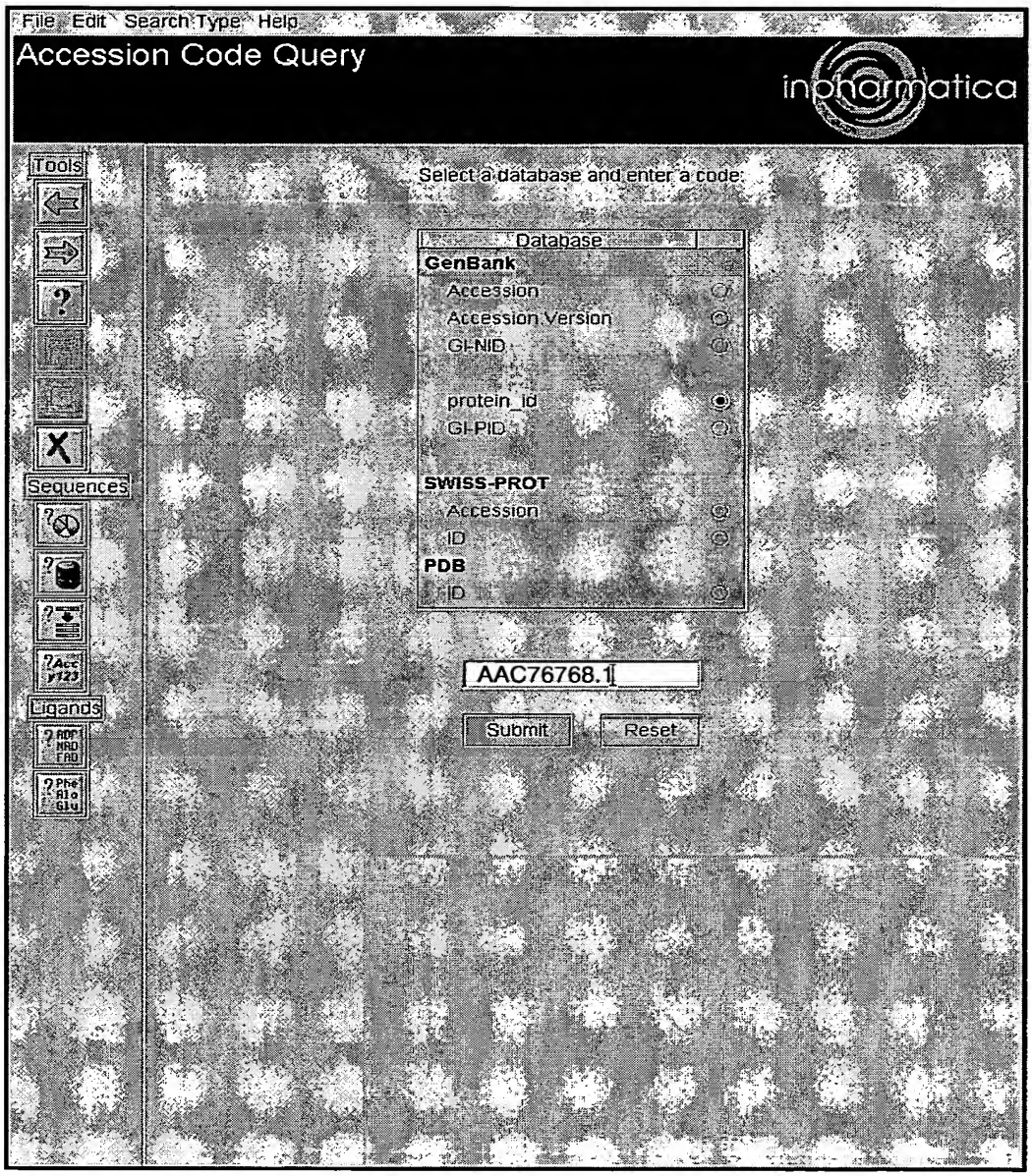
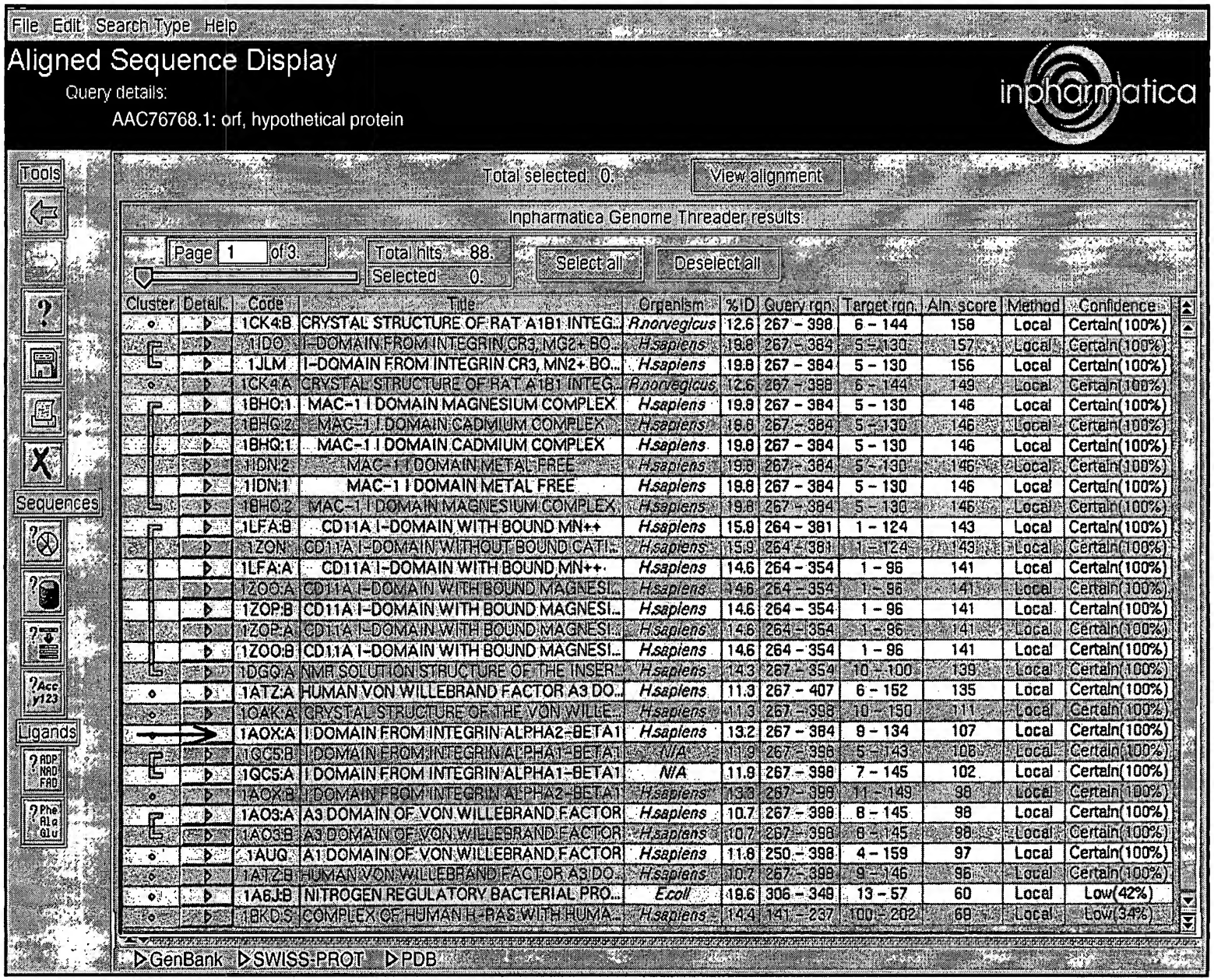
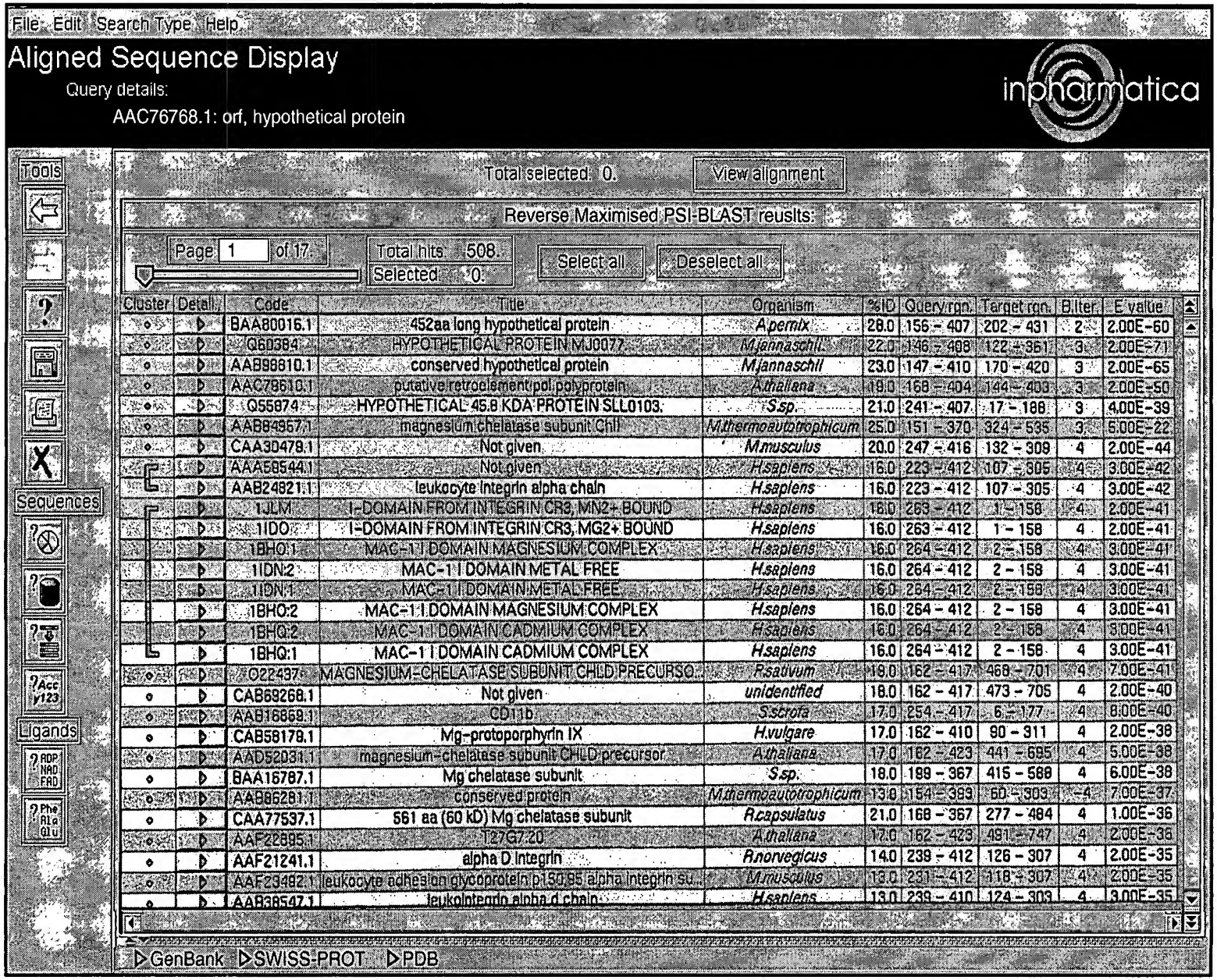

File Edit Search Type Help
Accession Code Query
inpharmatica
Tools
Sequences
Ligands
Select a database and enter a code:
Database
GenBank
Accession
Accession.Version
GI-NID
protein_id
GI-PID
SWISS-PROT
Accession
ID
PDB
ID
AAC76768.1
Submit
Reset

FIG. 14B




| Cluster | Detail | Code | Title | Organism | %ID | Query rgn. | Target rgn. | Aln. score | Method | Confidence |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 1CK4:B | CRYSTAL STRUCTURE OF RAT A1B1 INTEG... | *R.norvegicus* | 12.6 | 267 – 398 | 6 – 144 | 158 | Local | Certain(100%) |
| | | 1IDO | I-DOMAIN FROM INTEGRIN CR3, MG2+ BO... | *H.sapiens* | 19.8 | 267 – 384 | 5 – 130 | 157 | Local | Certain(100%) |
| | | 1JLM | I-DOMAIN FROM INTEGRIN CR3, MN2+ BO... | *H.sapiens* | 19.8 | 267 – 384 | 5 – 130 | 156 | Local | Certain(100%) |
| | | 1CK4:A | CRYSTAL STRUCTURE OF RAT A1B1 INTEG... | *R.norvegicus* | 12.6 | 267 – 398 | 6 – 144 | 149 | Local | Certain(100%) |
| | | 1BHQ:1 | MAC-1 I DOMAIN MAGNESIUM COMPLEX | *H.sapiens* | 19.8 | 267 – 384 | 5 – 130 | 146 | Local | Certain(100%) |
| | | 1BHQ:2 | MAC-1 I DOMAIN CADMIUM COMPLEX | *H.sapiens* | 19.8 | 267 – 384 | 5 – 130 | 146 | Local | Certain(100%) |
| | | 1BHQ:1 | MAC-1 I DOMAIN CADMIUM COMPLEX | *H.sapiens* | 19.8 | 267 – 384 | 5 – 130 | 146 | Local | Certain(100%) |
| | | 1IDN:2 | MAC-1 I DOMAIN METAL FREE | *H.sapiens* | 19.8 | 267 – 384 | 5 – 130 | 146 | Local | Certain(100%) |
| | | 1IDN:1 | MAC-1 I DOMAIN METAL FREE | *H.sapiens* | 19.8 | 267 – 384 | 5 – 130 | 146 | Local | Certain(100%) |
| | | 1BHQ:2 | MAC-1 I DOMAIN MAGNESIUM COMPLEX | *H.sapiens* | 19.8 | 267 – 384 | 5 – 130 | 146 | Local | Certain(100%) |
| | | 1LFA:B | CD11A I-DOMAIN WITH BOUND MN++ | *H.sapiens* | 15.8 | 264 – 381 | 1 – 124 | 143 | Local | Certain(100%) |
| | | 1ZON | CD11A I-DOMAIN WITHOUT BOUND CATI... | *H.sapiens* | 15.9 | 264 – 381 | 1 – 124 | 143 | Local | Certain(100%) |
| | | 1LFA:A | CD11A I-DOMAIN WITH BOUND MN++ | *H.sapiens* | 14.6 | 264 – 354 | 1 – 96 | 141 | Local | Certain(100%) |
| | | 1ZOO:A | CD11A I-DOMAIN WITH BOUND MAGNESI... | *H.sapiens* | 14.6 | 264 – 354 | 1 – 96 | 141 | Local | Certain(100%) |
| | | 1ZOP:B | CD11A I-DOMAIN WITH BOUND MAGNESI... | *H.sapiens* | 14.6 | 264 – 354 | 1 – 96 | 141 | Local | Certain(100%) |
| | | 1ZOP:A | CD11A I-DOMAIN WITH BOUND MAGNESI... | *H.sapiens* | 14.6 | 264 – 354 | 1 – 96 | 141 | Local | Certain(100%) |
| | | 1ZOO:B | CD11A I-DOMAIN WITH BOUND MAGNESI... | *H.sapiens* | 14.6 | 264 – 354 | 1 – 96 | 141 | Local | Certain(100%) |
| | | 1DGQ:A | NMR SOLUTION STRUCTURE OF THE INSER... | *H.sapiens* | 14.3 | 267 – 354 | 10 – 100 | 139 | Local | Certain(100%) |
| | | 1ATZ:A | HUMAN VON WILLEBRAND FACTOR A3 DO... | *H.sapiens* | 11.3 | 267 – 407 | 6 – 152 | 135 | Local | Certain(100%) |
| | | 1OAK:A | CRYSTAL STRUCTURE OF THE VON WILLE... | *H.sapiens* | 11.3 | 267 – 398 | 10 – 150 | 111 | Local | Certain(100%) |
| | | 1AOX:A | I DOMAIN FROM INTEGRIN ALPHA2-BETA1 | *H.sapiens* | 13.2 | 267 – 384 | 9 – 134 | 107 | Local | Certain(100%) |
| | | 1QC5:B | I DOMAIN FROM INTEGRIN ALPHA1-BETA1 | *N/A* | 11.9 | 267 – 398 | 5 – 143 | 106 | Local | Certain(100%) |
| | | 1QC5:A | I DOMAIN FROM INTEGRIN ALPHA1-BETA1 | *N/A* | 11.9 | 267 – 398 | 7 – 145 | 102 | Local | Certain(100%) |
| | | 1AOX:B | I DOMAIN FROM INTEGRIN ALPHA2-BETA1 | *H.sapiens* | 13.3 | 267 – 398 | 11 – 149 | 98 | Local | Certain(100%) |
| | | 1AO3:A | A3 DOMAIN OF VON WILLEBRAND FACTOR | *H.sapiens* | 10.7 | 267 – 398 | 8 – 145 | 98 | Local | Certain(100%) |
| | | 1AO3:B | A3 DOMAIN OF VON WILLEBRAND FACTOR | *H.sapiens* | 10.7 | 267 – 398 | 8 – 145 | 98 | Local | Certain(100%) |
| | | 1AUQ | A1 DOMAIN OF VON WILLEBRAND FACTOR | *H.sapiens* | 11.8 | 250 – 398 | 4 – 159 | 97 | Local | Certain(100%) |
| | | 1ATZ:B | HUMAN VON WILLEBRAND FACTOR A3 DO... | *H.sapiens* | 10.7 | 267 – 398 | 9 – 146 | 96 | Local | Certain(100%) |
| | | 1A6J:B | NITROGEN REGULATORY BACTERIAL PRO... | *E.coli* | 19.6 | 306 – 349 | 13 – 57 | 60 | Local | Low(42%) |
| | | 1BKD:S | COMPLEX OF HUMAN H-RAS WITH HUMA... | *H.sapiens* | 14.4 | 141 – 237 | 100 – 202 | 60 | Local | Low(34%) |

FIG. 14C


File Edit Search Type Help
Aligned Sequence Display
Query details:
AAC76768.1: orf, hypothetical protein
inpharmatica
Tools
Sequences
Ligands
Total selected: 0.
View alignment
Reverse Maximised PSI-BLAST reuslts:
Page 1 of 17.
Total hits: 508.
Selected: 0.
Select all
Deselect all
Cluster Detail Code Title Organism %ID Query rgn. Target rgn. Bit scr. E value
BAA80016.1 452aa long hypothetical protein A.pernix 28.0 156 - 407 202 - 431 2 2.00E-60
Q60384 HYPOTHETICAL PROTEIN MJ0077. M.jannaschii. 22.0 146 - 408 122 - 361 3 2.00E-71
AAB98810.1 conserved hypothetical protein M.jannaschii 23.0 147 - 410 170 - 420 3 2.00E-65
AAC78610.1 putative retroelement pol polyprotein A.thaliana 19.0 168 - 404 144 - 403 3 2.00E-50
Q55874 HYPOTHETICAL 45.8 KDA PROTEIN SLL0103. S.sp. 21.0 241 - 407 17 - 188 3 4.00E-39
AAB84957.1 magnesium chelatase subunit ChlI M.thermoautotrophicum 25.0 151 - 370 324 - 535 3 5.00E-22
CAA30479.1 Not given M.musculus 20.0 247 - 416 132 - 309 4 2.00E-44
AAA58544.1 Not given H.sapiens 16.0 223 - 412 107 - 305 4 3.00E-42
AAB24821.1 leukocyte integrin alpha chain H.sapiens 16.0 223 - 412 107 - 305 4 3.00E-42
1JLM I-DOMAIN FROM INTEGRIN CR3, MN2+ BOUND H.sapiens 16.0 263 - 412 1 - 158 4 2.00E-41
1IDO I-DOMAIN FROM INTEGRIN CR3, MG2+ BOUND H.sapiens 16.0 263 - 412 1 - 158 4 2.00E-41
1BHO:1 MAC-1 I DOMAIN MAGNESIUM COMPLEX H.sapiens 16.0 264 - 412 2 - 158 4 3.00E-41
1IDN:2 MAC-1 I DOMAIN METAL FREE H.sapiens 16.0 264 - 412 2 - 158 4 3.00E-41
1IDN:1 MAC-1 I DOMAIN METAL FREE H.sapiens 16.0 264 - 412 2 - 158 4 3.00E-41
1BHO:2 MAC-1 I DOMAIN MAGNESIUM COMPLEX H.sapiens 16.0 264 - 412 2 - 158 4 3.00E-41
1BHQ:2 MAC-1 I DOMAIN CADMIUM COMPLEX H.sapiens 16.0 264 - 412 2 - 158 4 3.00E-41
1BHQ:1 MAC-1 I DOMAIN CADMIUM COMPLEX H.sapiens 16.0 264 - 412 2 - 158 4 3.00E-41
O22437 MAGNESIUM-CHELATASE SUBUNIT CHLD PRECURSO... P.sativum 18.0 162 - 417 468 - 701 4 7.00E-41
CAB69268.1 Not given unidentified 18.0 162 - 417 473 - 705 4 2.00E-40
AAB16869.1 CD11b S.scrofa 17.0 254 - 417 6 - 177 4 8.00E-40
CAB58178.1 Mg-protoporphyrin IX H.vulgare 17.0 162 - 410 90 - 311 4 2.00E-38
AAD52031.1 magnesium-chelatase subunit CHLD precursor A.thaliana 17.0 162 - 423 441 - 695 4 5.00E-38
BAA16787.1 Mg chelatase subunit S.sp. 18.0 189 - 367 415 - 589 4 6.00E-38
AAB86281.1 conserved protein M.thermoautotrophicum 13.0 154 - 389 60 - 303 4 7.00E-37
CAA77537.1 561 aa (60 kD) Mg chelatase subunit R.capsulatus 21.0 168 - 367 277 - 484 4 1.00E-36
AAF22895.1 T27G7.20 A.thaliana 17.0 162 - 423 491 - 747 4 2.00E-36
AAF21241.1 alpha D integrin R.norvegicus 14.0 239 - 412 126 - 307 4 2.00E-35
AAF23482.1 leukocyte adhesion glycoprotein p150,95 alpha integrin su... M.musculus 13.0 231 - 412 118 - 307 4 2.00E-35
AAB38547.1 leukointegrin alpha d chain H.sapiens 13.0 239 - 410 124 - 303 4 3.00E-35
GenBank SWISS-PROT PDB

# FIG. 15

AlEye output (December 13, 2000 3:07 PM)

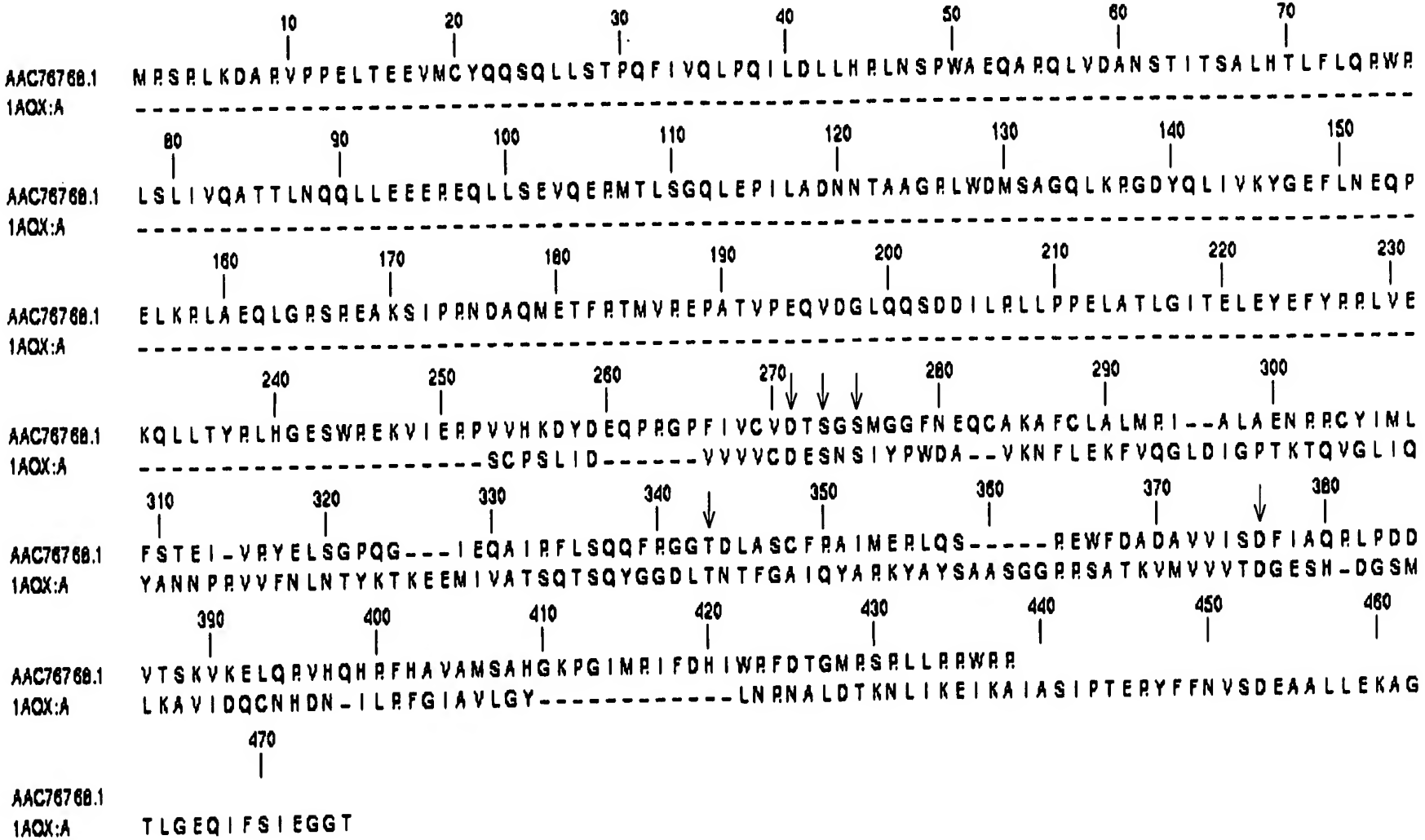


FIG. 16B

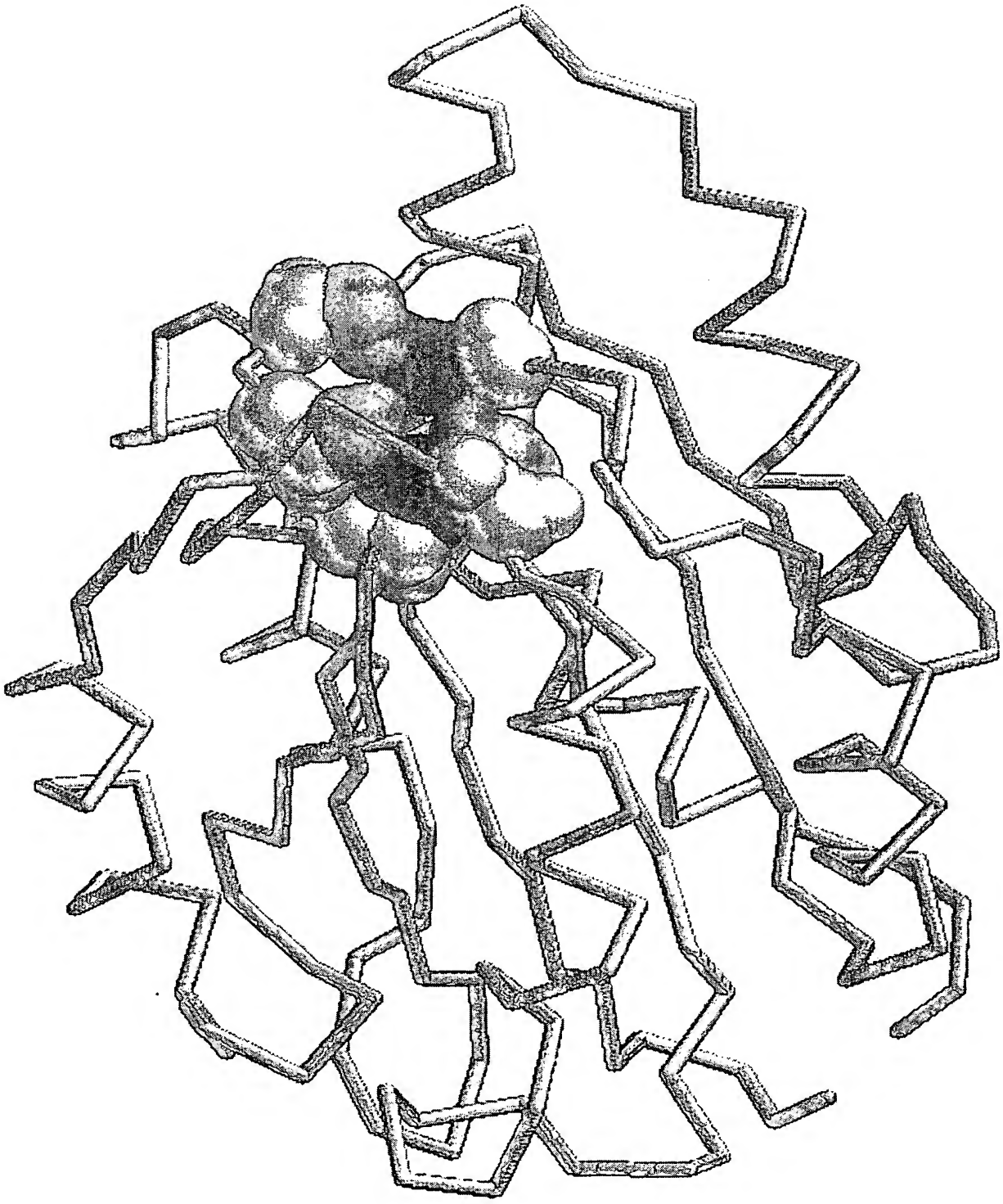
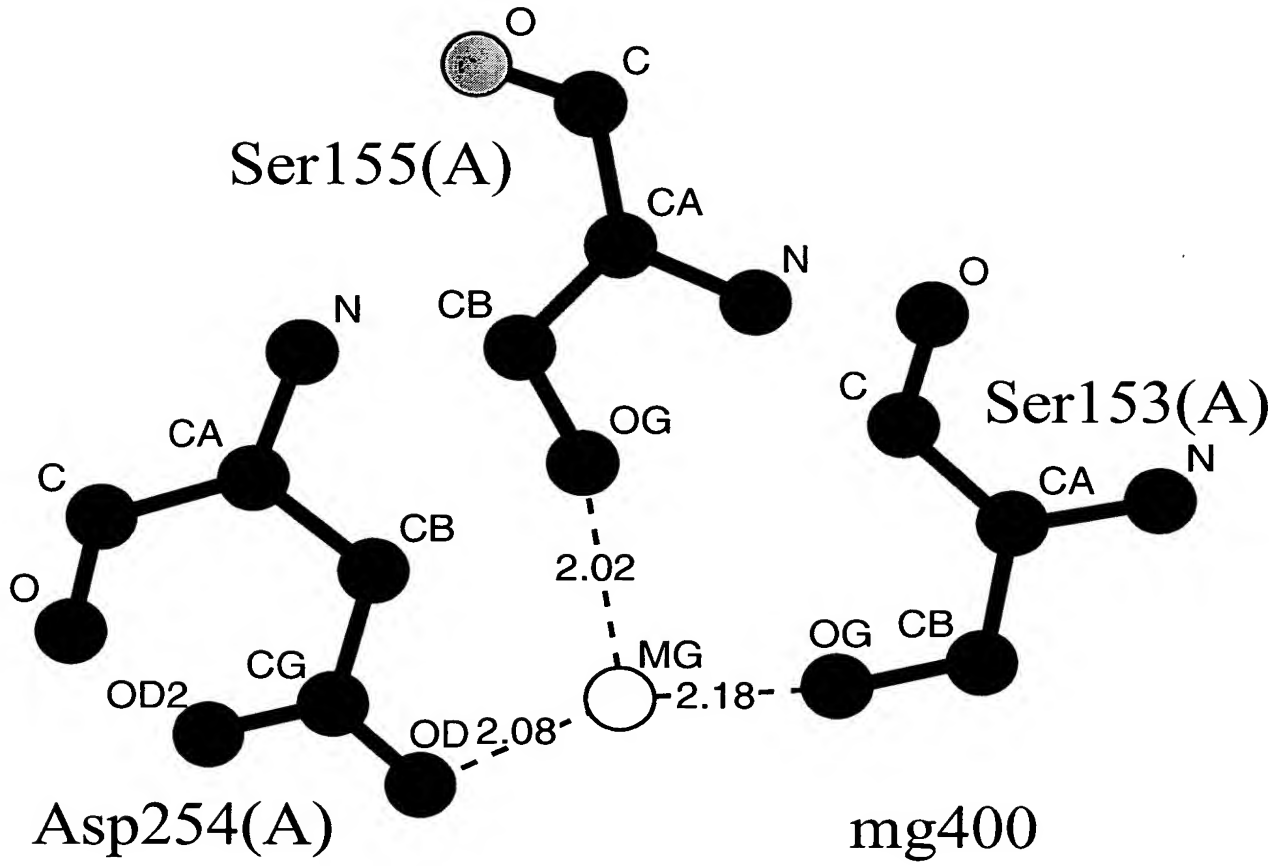


FIG. 16A



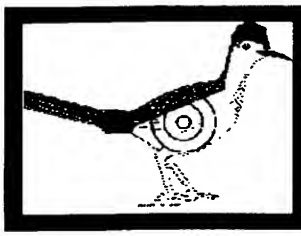

1aox: MG400

FIG. 17



File Edit View Go Communicator Help

Bookmarks Location http: //victoria. inpharmatica.co.uk/~volker/BPD3target.html



Target Mining Interface

inpharmatica

## Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): 1JLM and chain (e.g. B):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

## Select Database

Release: DEVF9 = BPD3

## Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

None

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

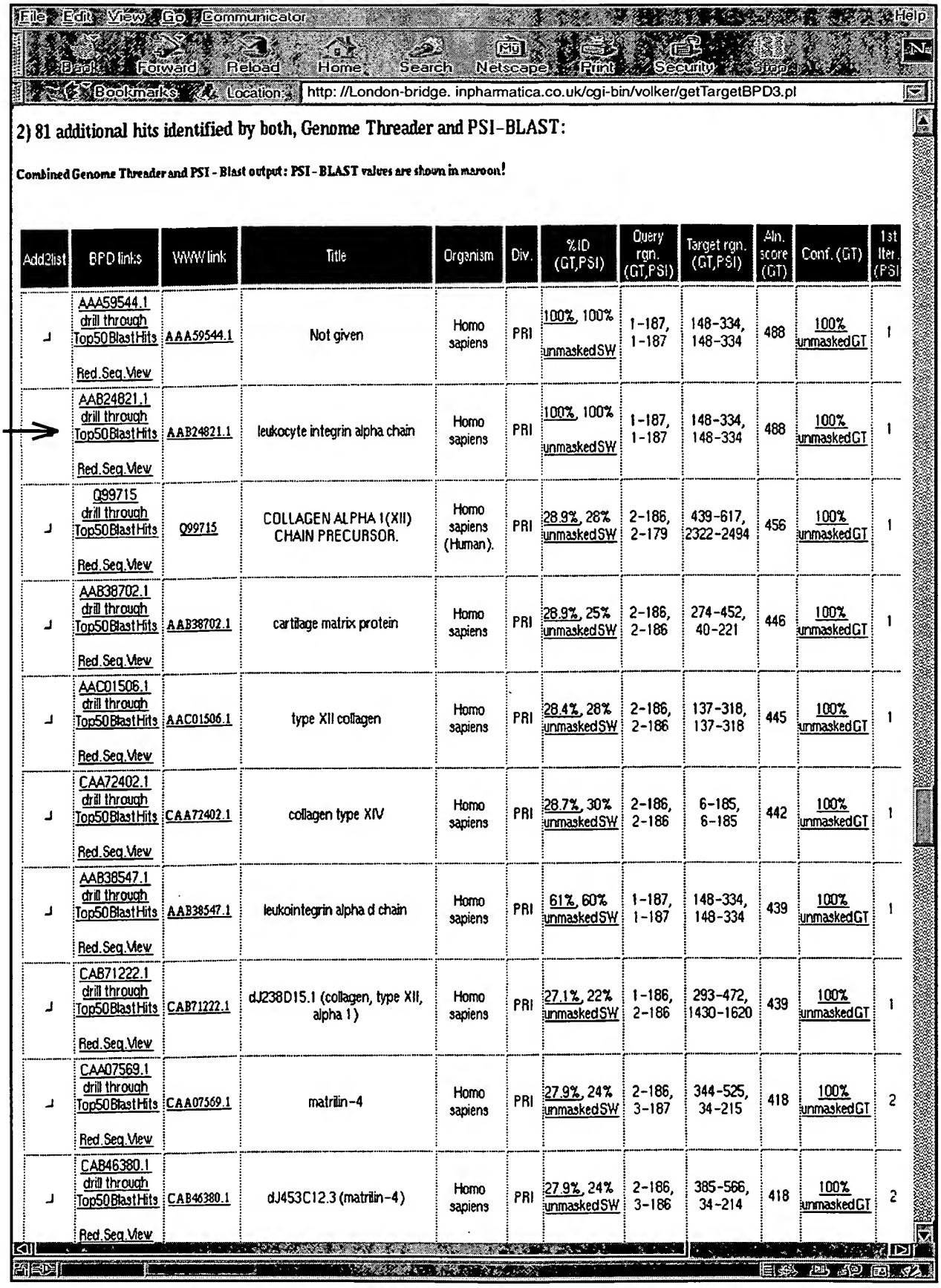
- Filter for the following SPECIES:

☐Homo sapiens ☐Rattus norvecicus (Rat) ☐Mus musculus (Mouse) ☐Danio rerio (Zebra fish)

100%

FIG. 18A



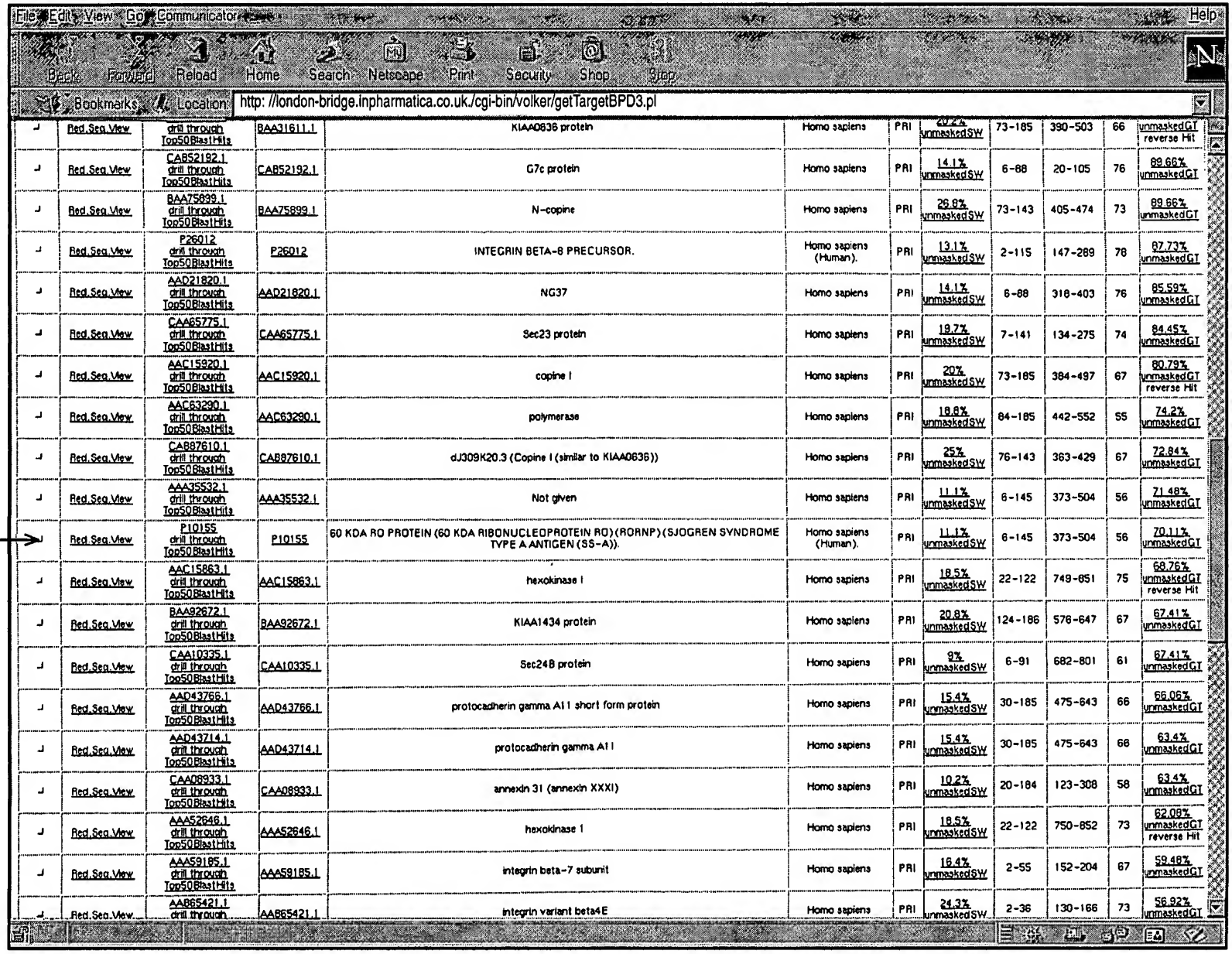



**2) 81 additional hits identified by both, Genome Threader and PSI-BLAST:**

**Combined Genome Threader and PSI - Blast output: PSI - BLAST values are shown in maroon!**

| Add2list | BPD links | WWW link | Title | Organism | Div. | %ID (GT,PSI) | Query rgn. (GT,PSI) | Target rgn. (GT,PSI) | Aln. score (GT) | Conf. (GT) | 1st Iter. (PSI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ⌟ | AAA59544.1 drill through Top50BlastHits Red.Seq.View | AAA59544.1 | Not given | Homo sapiens | PRI | 100%, 100% unmaskedSW | 1-187, 1-187 | 148-334, 148-334 | 488 | 100% unmaskedGT | 1 |
| → ⌟ | AAB24821.1 drill through Top50BlastHits Red.Seq.View | AAB24821.1 | leukocyte integrin alpha chain | Homo sapiens | PRI | 100%, 100% unmaskedSW | 1-187, 1-187 | 148-334, 148-334 | 488 | 100% unmaskedGT | 1 |
| ⌟ | Q99715 drill through Top50BlastHits Red.Seq.View | Q99715 | COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR. | Homo sapiens (Human). | PRI | 28.9%, 26% unmaskedSW | 2-186, 2-179 | 439-617, 2322-2494 | 456 | 100% unmaskedGT | 1 |
| ⌟ | AAB38702.1 drill through Top50BlastHits Red.Seq.View | AAB38702.1 | cartilage matrix protein | Homo sapiens | PRI | 28.9%, 25% unmaskedSW | 2-186, 2-186 | 274-452, 40-221 | 446 | 100% unmaskedGT | 1 |
| ⌟ | AAC01506.1 drill through Top50BlastHits Red.Seq.View | AAC01506.1 | type XII collagen | Homo sapiens | PRI | 28.4%, 28% unmaskedSW | 2-186, 2-186 | 137-318, 137-318 | 445 | 100% unmaskedGT | 1 |
| ⌟ | CAA72402.1 drill through Top50BlastHits Red.Seq.View | CAA72402.1 | collagen type XIV | Homo sapiens | PRI | 28.7%, 30% unmaskedSW | 2-186, 2-186 | 6-185, 6-185 | 442 | 100% unmaskedGT | 1 |
| ⌟ | AAB38547.1 drill through Top50BlastHits Red.Seq.View | AAB38547.1 | leukointegrin alpha d chain | Homo sapiens | PRI | 61%, 60% unmaskedSW | 1-187, 1-187 | 148-334, 148-334 | 439 | 100% unmaskedGT | 1 |
| ⌟ | CAB71222.1 drill through Top50BlastHits Red.Seq.View | CAB71222.1 | dJ238D15.1 (collagen, type XII, alpha 1) | Homo sapiens | PRI | 27.1%, 22% unmaskedSW | 1-186, 2-186 | 293-472, 1430-1620 | 439 | 100% unmaskedGT | 1 |
| ⌟ | CAA07569.1 drill through Top50BlastHits Red.Seq.View | CAA07569.1 | matrilin-4 | Homo sapiens | PRI | 27.9%, 24% unmaskedSW | 2-186, 3-187 | 344-525, 34-215 | 418 | 100% unmaskedGT | 2 |
| ⌟ | CAB46380.1 drill through Top50BlastHits Red.Seq.View | CAB46380.1 | dJ453C12.3 (matrilin-4) | Homo sapiens | PRI | 27.9%, 24% unmaskedSW | 2-186, 3-186 | 385-566, 34-214 | 418 | 100% unmaskedGT | 2 |

FIG. 18B




| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Red.Seq.View | drill through Top50BlastHits | BAA31611.1 | KIAA0636 protein | Homo sapiens | PRI | 40.2% unmaskedSW | 73-185 | 390-503 | 66 | unmaskedGI reverse Hit |
| | Red.Seq.View | CAB52192.1 drill through Top50BlastHits | CAB52192.1 | G7c protein | Homo sapiens | PRI | 14.1% unmaskedSW | 6-88 | 20-105 | 76 | 89.66% unmaskedGI |
| | Red.Seq.View | BAA75899.1 drill through Top50BlastHits | BAA75899.1 | N-copine | Homo sapiens | PRI | 26.8% unmaskedSW | 73-143 | 405-474 | 73 | 89.66% unmaskedGI |
| | Red.Seq.View | P26012 drill through Top50BlastHits | P26012 | INTEGRIN BETA-8 PRECURSOR. | Homo sapiens (Human). | PRI | 13.1% unmaskedSW | 2-115 | 147-269 | 78 | 87.73% unmaskedGI |
| | Red.Seq.View | AAD21820.1 drill through Top50BlastHits | AAD21820.1 | NG37 | Homo sapiens | PRI | 14.1% unmaskedSW | 6-88 | 318-403 | 76 | 85.59% unmaskedGI |
| | Red.Seq.View | CAA65775.1 drill through Top50BlastHits | CAA65775.1 | Sec23 protein | Homo sapiens | PRI | 19.7% unmaskedSW | 7-141 | 134-275 | 74 | 84.45% unmaskedGI |
| | Red.Seq.View | AAC15920.1 drill through Top50BlastHits | AAC15920.1 | copine I | Homo sapiens | PRI | 20% unmaskedSW | 73-185 | 384-497 | 67 | 80.79% unmaskedGI reverse Hit |
| | Red.Seq.View | AAC63290.1 drill through Top50BlastHits | AAC63290.1 | polymerase | Homo sapiens | PRI | 18.6% unmaskedSW | 84-185 | 442-552 | 55 | 74.2% unmaskedGI |
| | Red.Seq.View | CAB87610.1 drill through Top50BlastHits | CAB87610.1 | dJ309K20.3 (Copine I (similar to KIAA0636)) | Homo sapiens | PRI | 25% unmaskedSW | 76-143 | 363-429 | 67 | 72.84% unmaskedGI |
| | Red.Seq.View | AAA35532.1 drill through Top50BlastHits | AAA35532.1 | Not given | Homo sapiens | PRI | 11.1% unmaskedSW | 6-145 | 373-504 | 56 | 71.48% unmaskedGI |
| → | Red.Seq.View | P10155 drill through Top50BlastHits | P10155 | 60 KDA RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO)(RORNP)(SJOGREN SYNDROME TYPE A ANTIGEN (SS-A)). | Homo sapiens (Human). | PRI | 11.1% unmaskedSW | 6-145 | 373-504 | 56 | 70.11% unmaskedGI |
| | Red.Seq.View | AAC15863.1 drill through Top50BlastHits | AAC15863.1 | hexokinase I | Homo sapiens | PRI | 18.5% unmaskedSW | 22-122 | 749-851 | 75 | 68.76% unmaskedGI reverse Hit |
| | Red.Seq.View | BAA92672.1 drill through Top50BlastHits | BAA92672.1 | KIAA1434 protein | Homo sapiens | PRI | 20.8% unmaskedSW | 124-186 | 576-647 | 67 | 67.41% unmaskedGI |
| | Red.Seq.View | CAA10335.1 drill through Top50BlastHits | CAA10335.1 | Sec24B protein | Homo sapiens | PRI | 9% unmaskedSW | 6-91 | 682-801 | 61 | 67.41% unmaskedGI |
| | Red.Seq.View | AAD43766.1 drill through Top50BlastHits | AAD43766.1 | protocadherin gamma A11 short form protein | Homo sapiens | PRI | 15.4% unmaskedSW | 30-185 | 475-643 | 66 | 66.06% unmaskedGI |
| | Red.Seq.View | AAD43714.1 drill through Top50BlastHits | AAD43714.1 | protocadherin gamma A11 | Homo sapiens | PRI | 15.4% unmaskedSW | 30-185 | 475-643 | 66 | 63.4% unmaskedGI |
| | Red.Seq.View | CAA08933.1 drill through Top50BlastHits | CAA08933.1 | annexin 31 (annexin XXXI) | Homo sapiens | PRI | 10.2% unmaskedSW | 20-184 | 123-308 | 58 | 63.4% unmaskedGI |
| | Red.Seq.View | AAA52646.1 drill through Top50BlastHits | AAA52646.1 | hexokinase 1 | Homo sapiens | PRI | 18.5% unmaskedSW | 22-122 | 750-852 | 73 | 62.09% unmaskedGI reverse Hit |
| | Red.Seq.View | AAA59185.1 drill through Top50BlastHits | AAA59185.1 | integrin beta-7 subunit | Homo sapiens | PRI | 16.4% unmaskedSW | 2-55 | 152-204 | 67 | 59.48% unmaskedGI |
| | Red.Seq.View | AAB65421.1 drill through | AAB65421.1 | integrin variant beta4E | Homo sapiens | PRI | 24.3% unmaskedSW | 2-36 | 130-166 | 73 | 56.92% unmaskedGI |

FIG. 18C

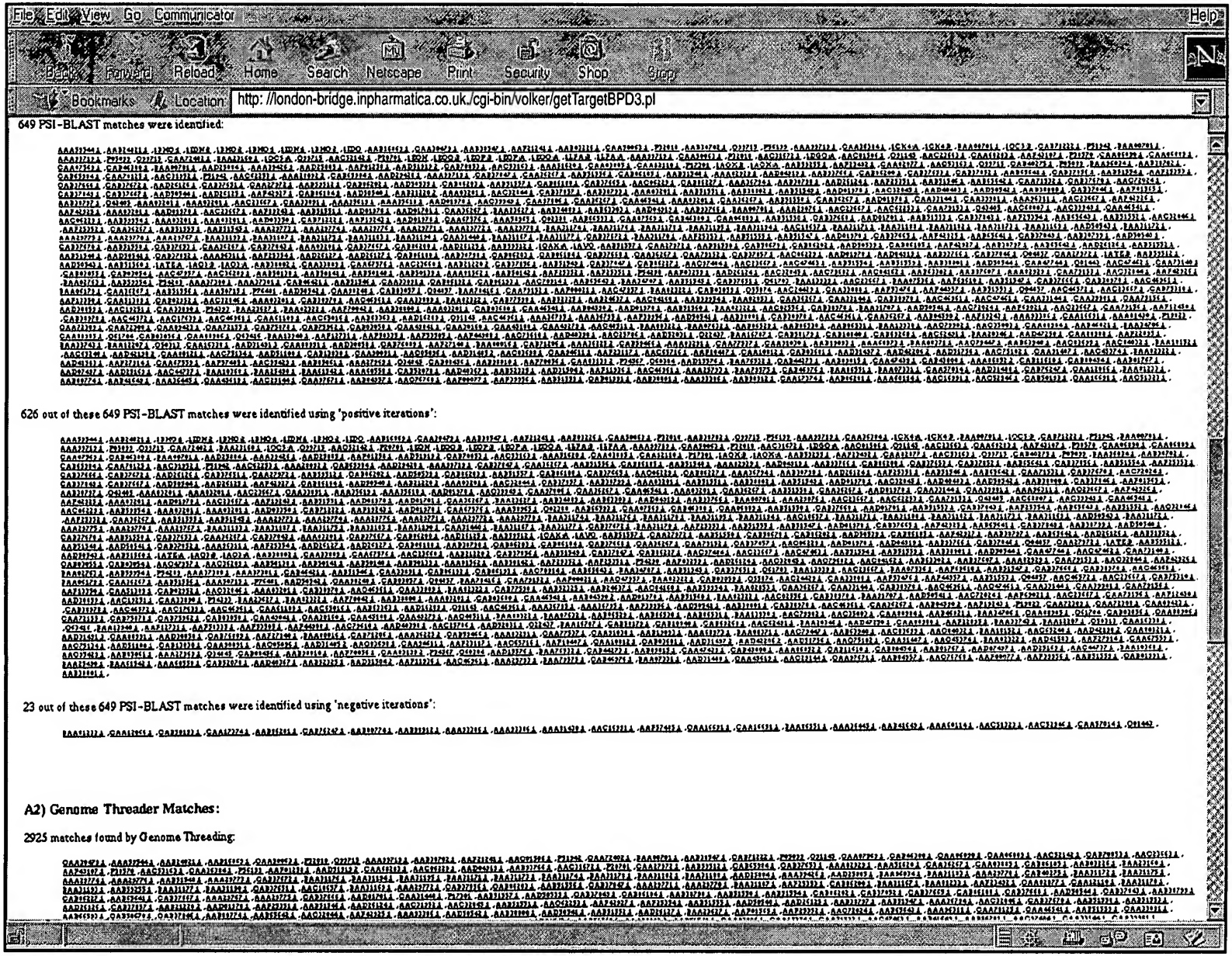

File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl
649 PSI-BLAST matches were identified:
626 out of these 649 PSI-BLAST matches were identified using 'positive iterations':
23 out of these 649 PSI-BLAST matches were identified using 'negative iterations':
A2) Genome Threader Matches:
2925 matches found by Genome Threading:

FIG. 19

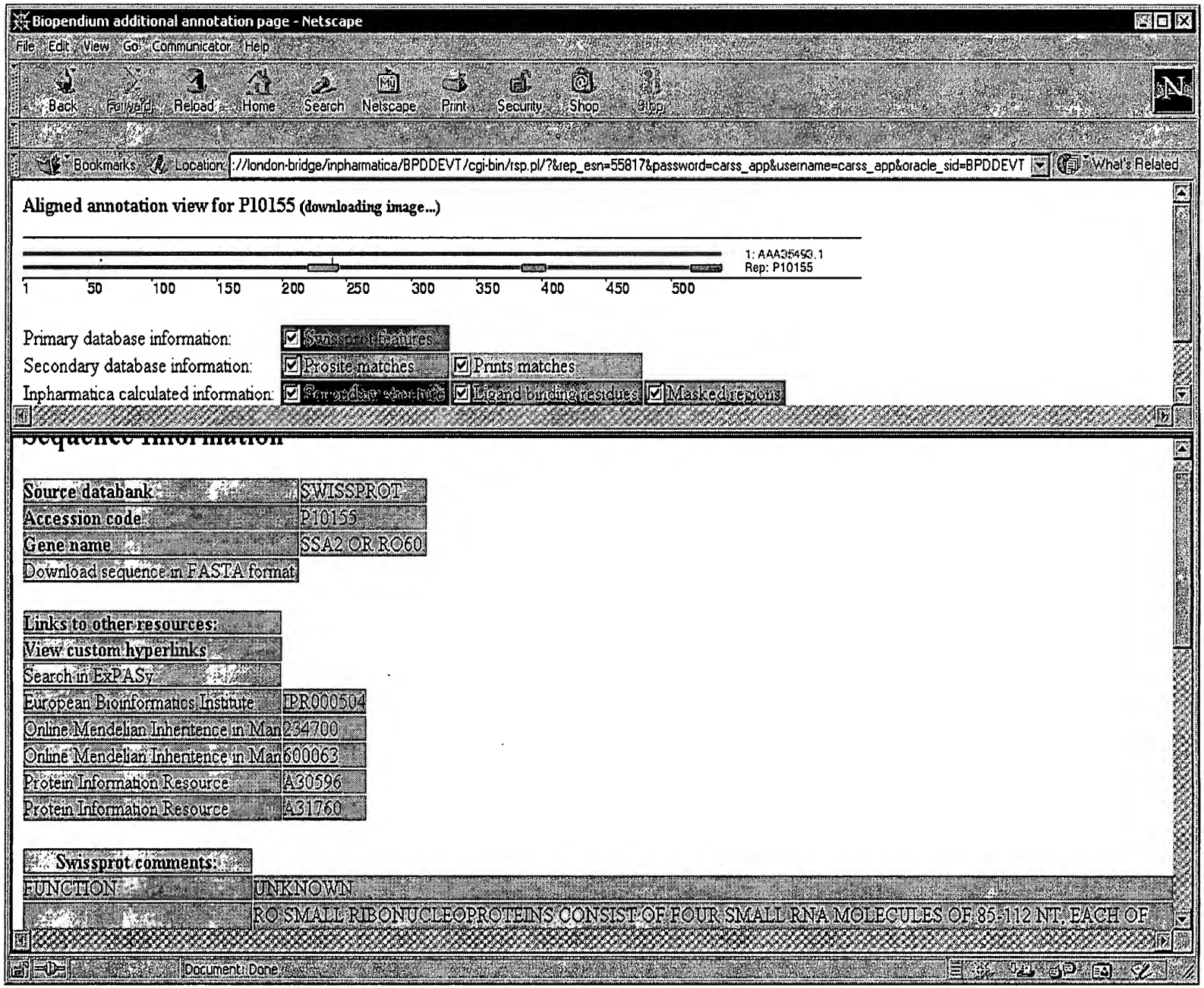

Biopendium additional annotation page - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: ://london-bridge/inpharmatica/BPDDEVT/cgi-bin/rsp.pl/?&rep_esn=55817&password=carss_app&username=carss_app&oracle_sid=BPDDEVT What's Related
Aligned annotation view for P10155 (downloading image...)
1: AAA35493.1
Rep: P10155
1 50 100 150 200 250 300 350 400 450 500
Primary database information:
Swissprot features
Secondary database information:
Prosite matches
Prints matches
Inpharmatica calculated information:
Secondary structure
Ligand binding residues
Masked regions
Source databank SWISSPROT
Accession code P10155
Gene name SSA2 OR RO60
Download sequence in FASTA format
Links to other resources:
View custom hyperlinks
Search in ExPASy
European Bioinformatics Institute PR000504
Online Mendelian Inheritence in Man 234700
Online Mendelian Inheritence in Man 600063
Protein Information Resource A30596
Protein Information Resource A31760
Swissprot comments:
FUNCTION UNKNOWN
RO SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT, EACH OF
Document: Done

FIG. 20

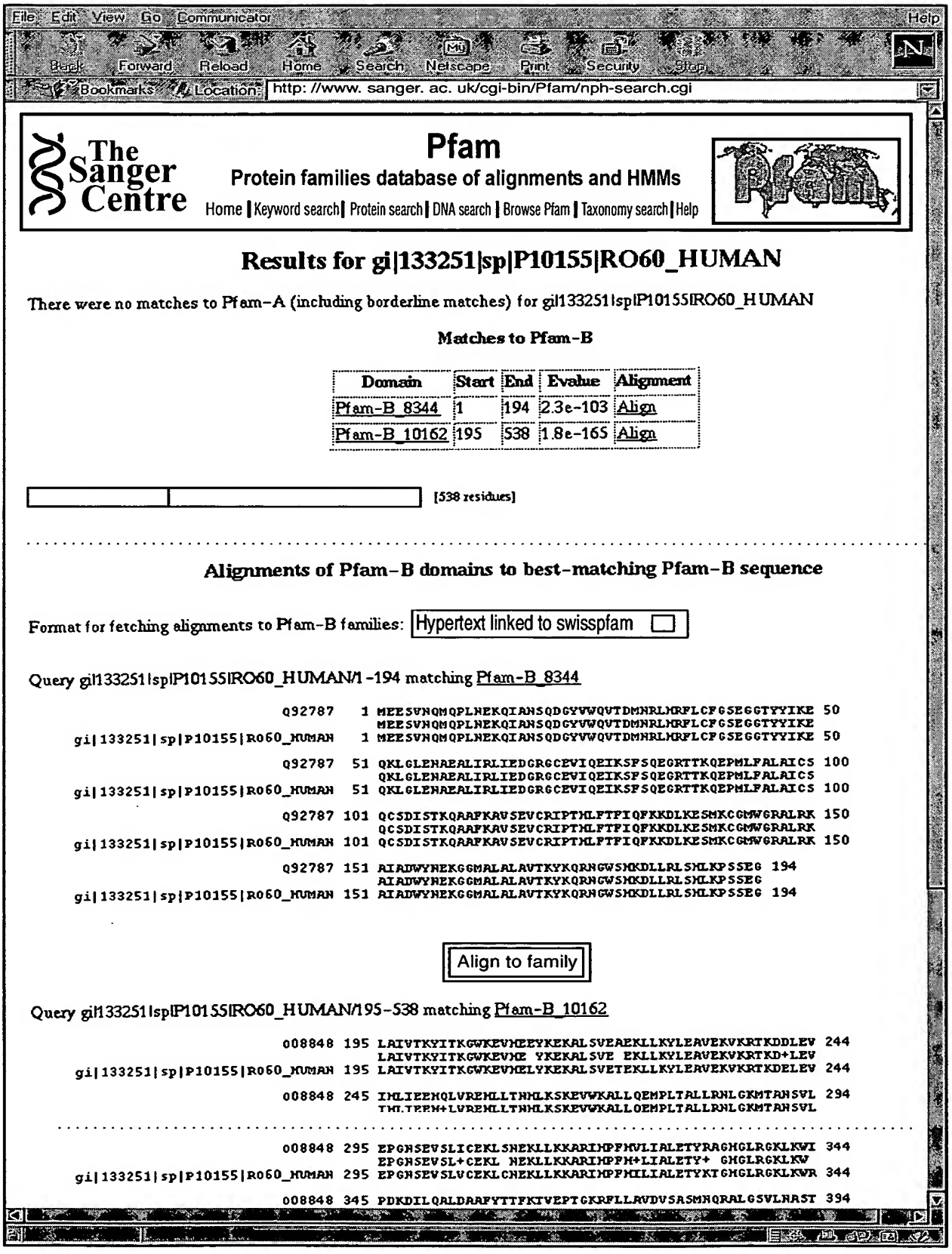
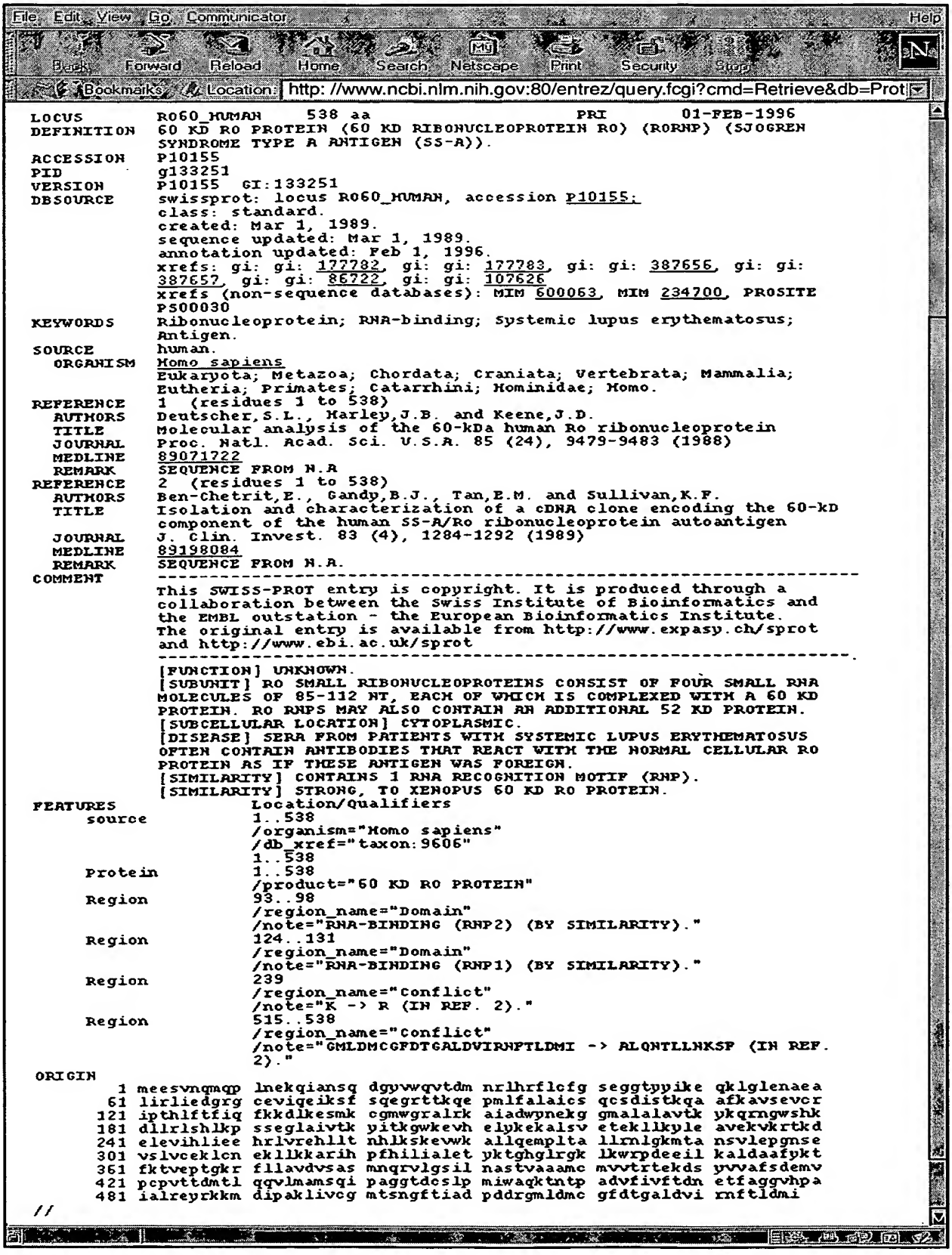

File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmarks Location: http: //www. sanger. ac. uk/cgi-bin/Pfam/nph-search.cgi
The Sanger Centre
Pfam
Protein families database of alignments and HMMs
Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help
Results for gi|133251|sp|P10155|RO60_HUMAN
There were no matches to Pfam-A (including borderline matches) for gi|133251|sp|P10155|RO60_HUMAN
Matches to Pfam-B
Domain Start End Evalue Alignment
Pfam-B_8344 1 194 2.3e-103 Align
Pfam-B_10162 195 538 1.8e-165 Align
[538 residues]
Alignments of Pfam-B domains to best-matching Pfam-B sequence
Format for fetching alignments to Pfam-B families: Hypertext linked to swissprot
Query gi|133251|sp|P10155|RO60_HUMAN/1-194 matching Pfam-B_8344
Q92787 1 MEESVNQMQPLNEKQIANSQDGYVWQVTDMNRLMRFLCFGSEGGTYYIKE 50
MEESVNQMQPLNEKQIANSQDGYVWQVTDMNRLMRFLCFGSEGGTYYIKE
gi|133251|sp|P10155|RO60_HUMAN 1 MEESVNQMQPLNEKQIANSQDGYVWQVTDMNRLMRFLCFGSEGGTYYIKE 50
Q92787 51 QKLGLENAEALIRLIEDGRGCEVIQEIKSFSQEGRTTKQEPMLFALAICS 100
QKLGLENAEALIRLIEDGRGCEVIQEIKSFSQEGRTTKQEPMLFALAICS
gi|133251|sp|P10155|RO60_HUMAN 51 QKLGLENAEALIRLIEDGRGCEVIQEIKSFSQEGRTTKQEPMLFALAICS 100
Q92787 101 QCSDISTKQAAFKAVSEVCRIPTHLFTFIQFKKDLKESMKCGMWGRALRK 150
QCSDISTKQAAFKAVSEVCRIPTHLFTFIQFKKDLKESMKCGMWGRALRK
gi|133251|sp|P10155|RO60_HUMAN 101 QCSDISTKQAAFKAVSEVCRIPTHLFTFIQFKKDLKESMKCGMWGRALRK 150
Q92787 151 RIADWYNEKGGMALALAVTKYKQRNGWSMKDLLRLSMLKPSSEG 194
RIADWYNEKGGMALALAVTKYKQRNGWSMKDLLRLSMLKPSSEG
gi|133251|sp|P10155|RO60_HUMAN 151 RIADWYNEKGGMALALAVTKYKQRNGWSMKDLLRLSMLKPSSEG 194
Align to family
Query gi|133251|sp|P10155|RO60_HUMAN/195-538 matching Pfam-B_10162
O08848 195 LAIVTKYITKGWKEVMEEYKEKALSVEAEKLLKYLEAVEKVKRTKDDLEV 244
LAIVTKYITKGWKEVME YKEKALSVE EKLLKYLEAVEKVKRTKD+LEV
gi|133251|sp|P10155|RO60_HUMAN 195 LAIVTKYITKGWKEVMELYKEKALSVETEKLLKYLEAVEKVKRTKDELEV 244
O08848 245 IMLIEEHQLVREMLLTHMLKSKEVWKALLQEMPLTALLRNLGKMTANSVL 294
IML EEH+LVREMLLTHMLKSKEVWKALLQEMPLTALLRNLGKMTANSVL
O08848 295 EPGNSEVSLICEKLSNEKLLKKARIHPFMVLIALETYRAGMGLRGKLKWI 344
EPGNSEVSL+CEKL NEKLLKKARIHPFM+LIALETY+ GMGLRGKLKW
gi|133251|sp|P10155|RO60_HUMAN 295 EPGNSEVSLVCEKLCNEKLLKKARIHPFMILIALETYKTGMGLRGKLKWR 344
O08848 345 PDKDILQALDRAFYTTFKIVEPTGKRFLLAVDVSASMNQRALGSVLNAST 394

FIG. 21



File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location: http: //www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=Prot

```
LOCUS       RO60_HUMAN     538 aa                    PRI       01-FEB-1996
DEFINITION  60 KD RO PROTEIN (60 KD RIBONUCLEOPROTEIN RO) (RORNP) (SJOGREN
            SYNDROME TYPE A ANTIGEN (SS-A)).
ACCESSION   P10155
PID         g133251
VERSION     P10155  GI:133251
DBSOURCE    swissprot: locus RO60_HUMAN, accession P10155;
            class: standard.
            created: Mar 1, 1989.
            sequence updated: Mar 1, 1989.
            annotation updated: Feb 1, 1996.
            xrefs: gi: gi: 177782, gi: gi: 177783, gi: gi: 387656, gi: gi:
            387657, gi: gi: 86722, gi: gi: 107626
            xrefs (non-sequence databases): MIM 600063, MIM 234700, PROSITE
            PS00030
KEYWORDS    Ribonucleoprotein; RNA-binding; Systemic lupus erythematosus;
            Antigen.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 538)
  AUTHORS   Deutscher,S.L., Harley,J.B. and Keene,J.D.
  TITLE     Molecular analysis of the 60-kDa human Ro ribonucleoprotein
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85 (24), 9479-9483 (1988)
  MEDLINE   89071722
  REMARK    SEQUENCE FROM N.A
REFERENCE   2  (residues 1 to 538)
  AUTHORS   Ben-Chetrit,E., Gandy,B.J., Tan,E.M. and Sullivan,K.F.
  TITLE     Isolation and characterization of a cDNA clone encoding the 60-kD
            component of the human SS-A/Ro ribonucleoprotein autoantigen
  JOURNAL   J. Clin. Invest. 83 (4), 1284-1292 (1989)
  MEDLINE   89198084
  REMARK    SEQUENCE FROM N.A.
COMMENT     ---------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------.
            [FUNCTION] UNKNOWN.
            [SUBUNIT] RO SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA
            MOLECULES OF 85-112 NT, EACH OF WHICH IS COMPLEXED WITH A 60 KD
            PROTEIN. RO RNPS MAY ALSO CONTAIN AN ADDITIONAL 52 KD PROTEIN.
            [SUBCELLULAR LOCATION] CYTOPLASMIC.
            [DISEASE] SERA FROM PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS
            OFTEN CONTAIN ANTIBODIES THAT REACT WITH THE NORMAL CELLULAR RO
            PROTEIN AS IF THESE ANTIGEN WAS FOREIGN.
            [SIMILARITY] CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
            [SIMILARITY] STRONG, TO XENOPUS 60 KD RO PROTEIN.
FEATURES                Location/Qualifiers
     source             1..538
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        1..538
     Protein            1..538
                        /product="60 KD RO PROTEIN"
     Region             93..98
                        /region_name="Domain"
                        /note="RNA-BINDING (RNP2) (BY SIMILARITY)."
     Region             124..131
                        /region_name="Domain"
                        /note="RNA-BINDING (RNP1) (BY SIMILARITY)."
     Region             239
                        /region_name="Conflict"
                        /note="K -> R (IN REF. 2)."
     Region             515..538
                        /region_name="Conflict"
                        /note="GMLDMCGFDTGALDVIRNPTLDMI -> ALQNTLLNKSP (IN REF.
                        2)."
ORIGIN
        1 meesvnqmqp lnekqiansq dgyvwqvtdm nrlhrflcfg seggtpyike qklglenaea
       61 lirliedgrg ceviqeiksf sqegrttkqe pmlfalaics qcsdistkqa afkavsevcr
      121 ipthlftfiq fkkdlkesmk cgmwgralrk aiadwpnekg gmalalavtk ykqrngwshk
      181 dllrlshlkp sseglaivtk yitkgwkevh elykekalsv eteklllkyle avekvkrtkd
      241 elevihliee hrlvrehllt nhlkskewwk allqemplta llrnlgkmta nsvlepgnse
      301 vslvceklcn eklikkarih pfhilialet yktghglrgk lkwrpdeeil kaldaafykt
      361 fktveptgkr fllavdvsas mnqrvlgsil nastvaaamc mvvtrtekds yvvafsdemv
      421 pcpvttdmtl qqvlmamsqi paggtdcslp miwaqktntp advfivftdn etfaggvhpa
      481 ialreyrkkm dipaklivcg mtsngftiad pddrgmldmc gfdtgaldvi rnftldmi
//
```

FIG. 22A

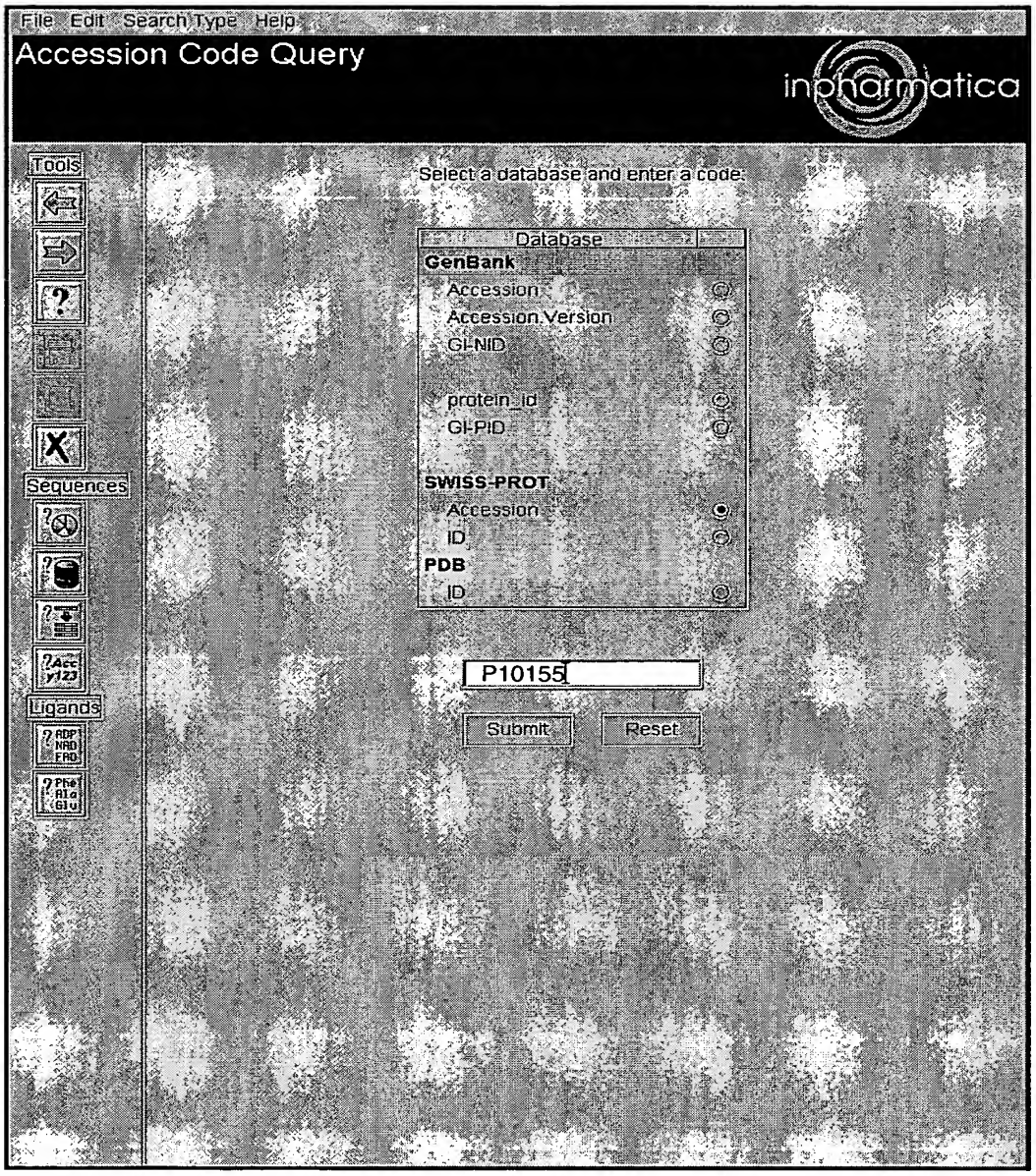
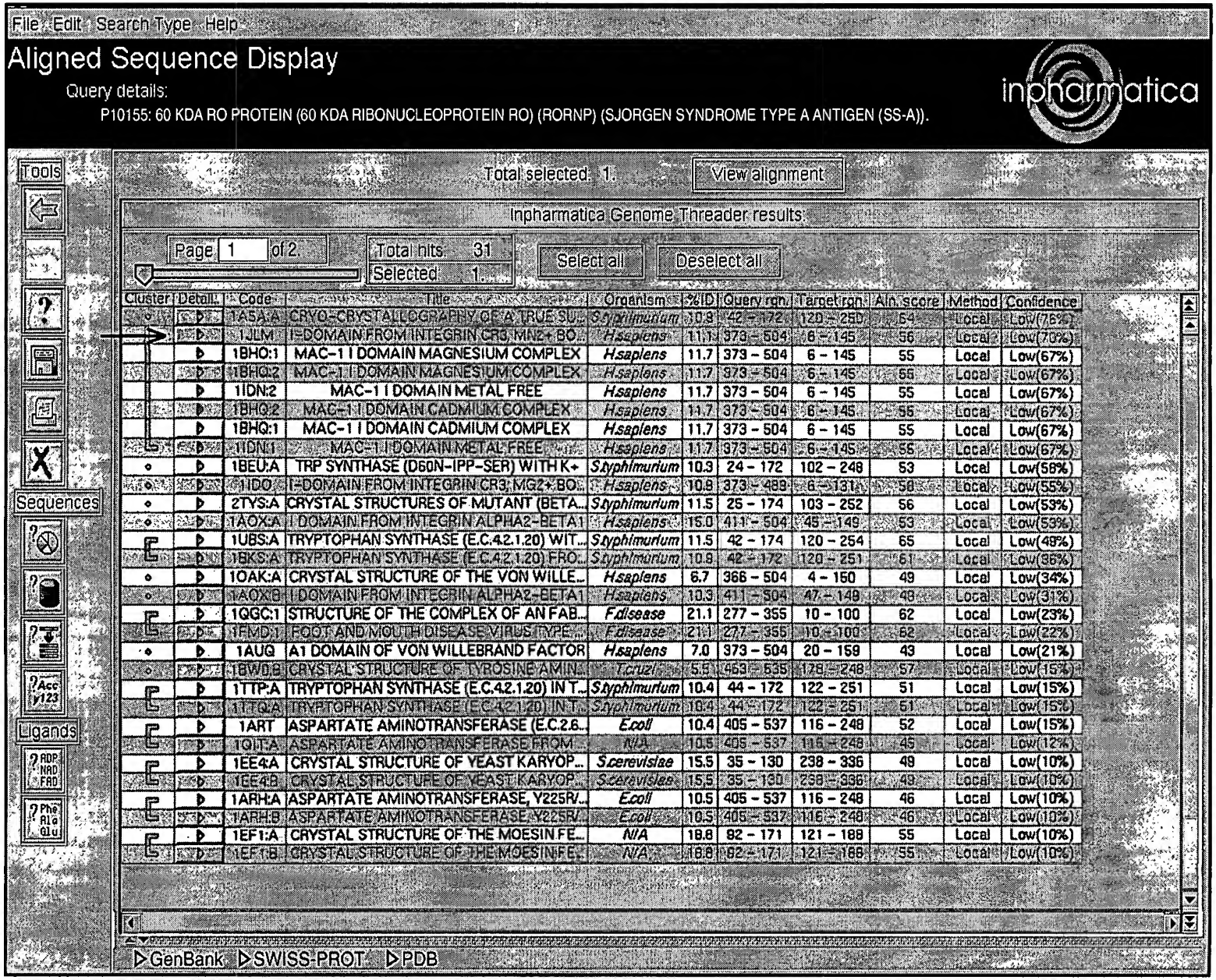
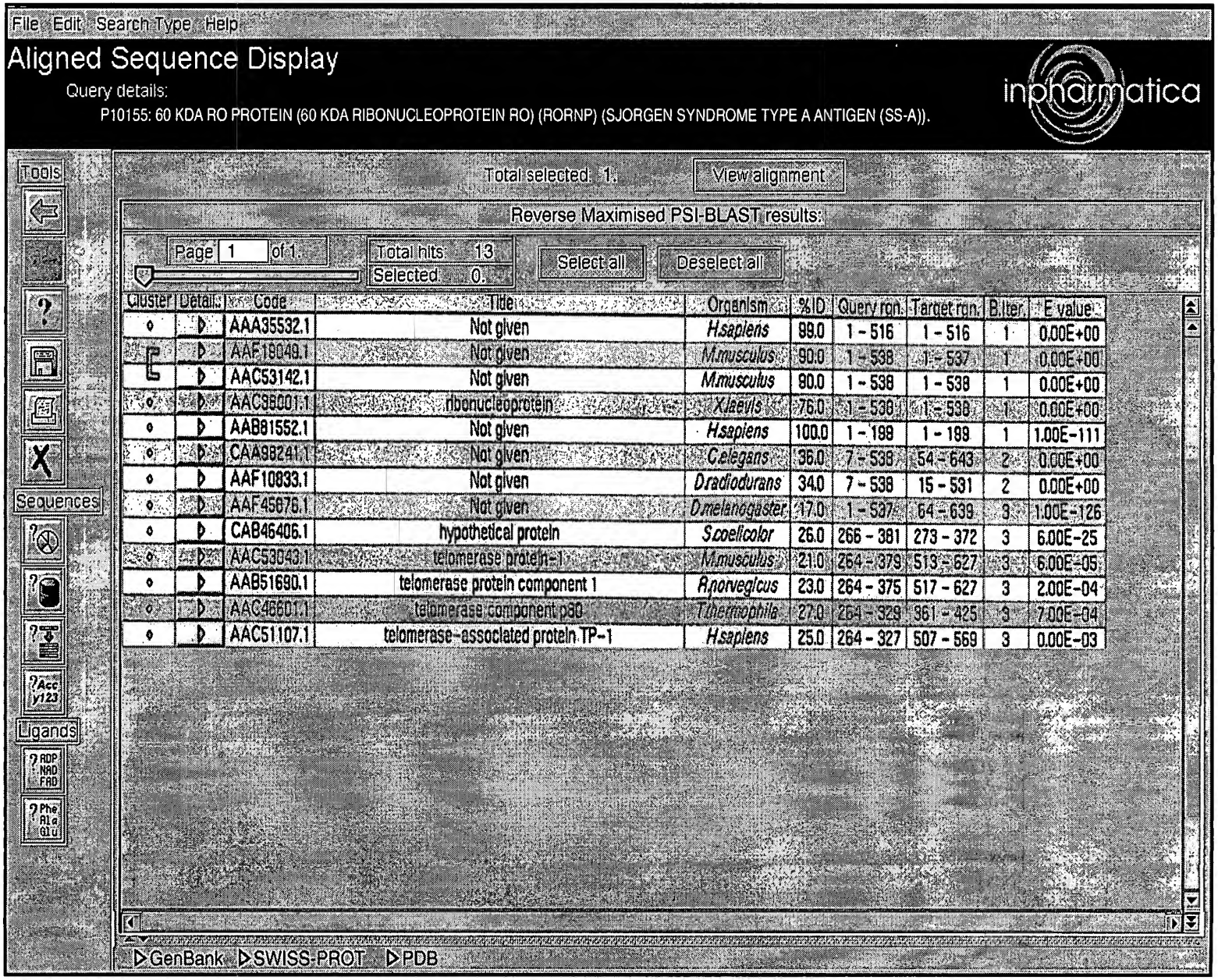

File Edit Search Type Help
Accession Code Query
inpharmatica
Tools
Sequences
Ligands
Select a database and enter a code:
Database
GenBank
Accession
Accession.Version
GI-NID
protein_id
GI-PID
SWISS-PROT
Accession
ID
PDB
ID
P10155
Submit
Reset

FIG. 22B




| Cluster | Detail | Code | Title | Organism | %ID | Query rgn. | Target rgn. | Aln. score | Method | Confidence |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 1ASA:A | CRYO-CRYSTALLOGRAPHY OF A TRUE SU... | *S.typhimurium* | 10.8 | 42 – 172 | 120 – 250 | 64 | Local | Low(76%) |
| | | 1JLM | I-DOMAIN FROM INTEGRIN CR3, MN2+ BO... | *H.sapiens* | 11.1 | 373 – 504 | 6 – 145 | 56 | Local | Low(70%) |
| | | 1BHO:1 | MAC-1 I DOMAIN MAGNESIUM COMPLEX | *H.sapiens* | 11.7 | 373 – 504 | 6 – 145 | 55 | Local | Low(67%) |
| | | 1BHQ:2 | MAC-1 I DOMAIN MAGNESIUM COMPLEX | *H.sapiens* | 11.7 | 373 – 504 | 6 – 145 | 55 | Local | Low(67%) |
| | | 1IDN:2 | MAC-1 I DOMAIN METAL FREE | *H.sapiens* | 11.7 | 373 – 504 | 6 – 145 | 55 | Local | Low(67%) |
| | | 1BHQ:2 | MAC-1 I DOMAIN CADMIUM COMPLEX | *H.sapiens* | 11.7 | 373 – 504 | 6 – 145 | 55 | Local | Low(67%) |
| | | 1BHQ:1 | MAC-1 I DOMAIN CADMIUM COMPLEX | *H.sapiens* | 11.7 | 373 – 504 | 6 – 145 | 55 | Local | Low(67%) |
| | | 1IDN:1 | MAC-1 I DOMAIN METAL FREE | *H.sapiens* | 11.7 | 373 – 504 | 6 – 145 | 55 | Local | Low(67%) |
| | | 1BEU:A | TRP SYNTHASE (D60N-IPP-SER) WITH K+ | *S.typhimurium* | 10.3 | 24 – 172 | 102 – 248 | 53 | Local | Low(58%) |
| | | 1IDO | I-DOMAIN FROM INTEGRIN CR3, MG2+ BO... | *H.sapiens* | 10.8 | 373 – 489 | 6 – 131 | 50 | Local | Low(55%) |
| | | 2TYS:A | CRYSTAL STRUCTURES OF MUTANT (BETA... | *S.typhimurium* | 11.5 | 25 – 174 | 103 – 252 | 56 | Local | Low(53%) |
| | | 1AOX:A | I DOMAIN FROM INTEGRIN ALPHA2-BETA1 | *H.sapiens* | 15.0 | 411 – 504 | 45 – 149 | 53 | Local | Low(53%) |
| | | 1UBS:A | TRYPTOPHAN SYNTHASE (E.C.4.2.1.20) WIT... | *S.typhimurium* | 11.5 | 42 – 174 | 120 – 254 | 65 | Local | Low(48%) |
| | | 1BKS:A | TRYPTOPHAN SYNTHASE (E.C.4.2.1.20) FRO... | *S.typhimurium* | 10.8 | 42 – 172 | 120 – 251 | 61 | Local | Low(36%) |
| | | 1OAK:A | CRYSTAL STRUCTURE OF THE VON WILLE... | *H.sapiens* | 6.7 | 366 – 504 | 4 – 150 | 49 | Local | Low(34%) |
| | | 1AOX:B | I DOMAIN FROM INTEGRIN ALPHA2-BETA1 | *H.sapiens* | 10.3 | 411 – 504 | 47 – 149 | 49 | Local | Low(31%) |
| | | 1QGC:1 | STRUCTURE OF THE COMPLEX OF AN FAB... | *F.disease* | 21.1 | 277 – 355 | 10 – 100 | 62 | Local | Low(23%) |
| | | 1FMD:1 | FOOT AND MOUTH DISEASE VIRUS TYPE... | *F.disease* | 21.1 | 277 – 355 | 10 – 100 | 62 | Local | Low(22%) |
| | | 1AUQ | A1 DOMAIN OF VON WILLEBRAND FACTOR | *H.sapiens* | 7.0 | 373 – 504 | 20 – 159 | 43 | Local | Low(21%) |
| | | 1BW0:B | CRYSTAL STRUCTURE OF TYROSINE AMIN... | *T.cruzi* | 5.5 | 463 – 535 | 178 – 248 | 57 | Local | Low(15%) |
| | | 1TTP:A | TRYPTOPHAN SYNTHASE (E.C.4.2.1.20) IN T... | *S.typhimurium* | 10.4 | 44 – 172 | 122 – 251 | 51 | Local | Low(15%) |
| | | 1TTQ:A | TRYPTOPHAN SYNTHASE (E.C.4.2.1.20) IN T... | *S.typhimurium* | 10.4 | 44 – 172 | 122 – 251 | 51 | Local | Low(15%) |
| | | 1ART | ASPARTATE AMINOTRANSFERASE (E.C.2.6... | *E.coli* | 10.4 | 405 – 537 | 116 – 248 | 52 | Local | Low(15%) |
| | | 1QLT:A | ASPARTATE AMINOTRANSFERASE FROM... | *N/A* | 10.5 | 405 – 537 | 116 – 248 | 45 | Local | Low(12%) |
| | | 1EE4:A | CRYSTAL STRUCTURE OF YEAST KARYOP... | *S.cerevisiae* | 15.5 | 35 – 130 | 238 – 336 | 49 | Local | Low(10%) |
| | | 1EE4:B | CRYSTAL STRUCTURE OF YEAST KARYOP... | *S.cerevisiae* | 15.5 | 35 – 130 | 238 – 336 | 49 | Local | Low(10%) |
| | | 1ARH:A | ASPARTATE AMINOTRANSFERASE, Y225R/... | *E.coli* | 10.5 | 405 – 537 | 116 – 248 | 46 | Local | Low(10%) |
| | | 1ARH:B | ASPARTATE AMINOTRANSFERASE, Y225R/... | *E.coli* | 10.5 | 405 – 537 | 116 – 248 | 46 | Local | Low(10%) |
| | | 1EF1:A | CRYSTAL STRUCTURE OF THE MOESIN FE... | *N/A* | 18.8 | 92 – 171 | 121 – 188 | 55 | Local | Low(10%) |
| | | 1EF1:B | CRYSTAL STRUCTURE OF THE MOESIN FE... | *N/A* | 18.8 | 92 – 171 | 121 – 188 | 55 | Local | Low(10%) |

FIG. 22B




| Cluster | Detail | Code | Title | Organism | %ID | Query rgn. | Target rgn. | Iter. | E value |
|---|---|---|---|---|---|---|---|---|---|
| | | AAA35532.1 | Not given | *H.sapiens* | 99.0 | 1 - 516 | 1 - 516 | 1 | 0.00E+00 |
| | | AAF18049.1 | Not given | *M.musculus* | 90.0 | 1 - 538 | 1 - 537 | 1 | 0.00E+00 |
| | | AAC53142.1 | Not given | *M.musculus* | 90.0 | 1 - 538 | 1 - 538 | 1 | 0.00E+00 |
| | | AAC98001.1 | ribonucleoprotein | *X.laevis* | 76.0 | 1 - 538 | 1 - 538 | 1 | 0.00E+00 |
| | | AAB81552.1 | Not given | *H.sapiens* | 100.0 | 1 - 199 | 1 - 199 | 1 | 1.00E-111 |
| | | CAA98241.1 | Not given | *C.elegans* | 36.0 | 7 - 538 | 54 - 643 | 2 | 0.00E+00 |
| | | AAF10833.1 | Not given | *D.radiodurans* | 34.0 | 7 - 538 | 15 - 531 | 2 | 0.00E+00 |
| | | AAF45676.1 | Not given | *D.melanogaster* | 17.0 | 1 - 537 | 64 - 639 | 3 | 1.00E-126 |
| | | CAB46406.1 | hypothetical protein | *S.coelicolor* | 26.0 | 266 - 381 | 273 - 372 | 3 | 6.00E-25 |
| | | AAC53043.1 | telomerase protein-1 | *M.musculus* | 21.0 | 264 - 379 | 513 - 627 | 3 | 6.00E-05 |
| | | AAB51690.1 | telomerase protein component 1 | *R.norvegicus* | 23.0 | 264 - 375 | 517 - 627 | 3 | 2.00E-04 |
| | | AAC46601.1 | telomerase component p80 | *T.thermophila* | 27.0 | 264 - 329 | 361 - 425 | 3 | 7.00E-04 |
| | | AAC51107.1 | telomerase-associated protein TP-1 | *H.sapiens* | 25.0 | 264 - 327 | 507 - 569 | 3 | 0.00E-03 |

# FIG. 23

AlEye output (January 2, 2001 1:29 PM)

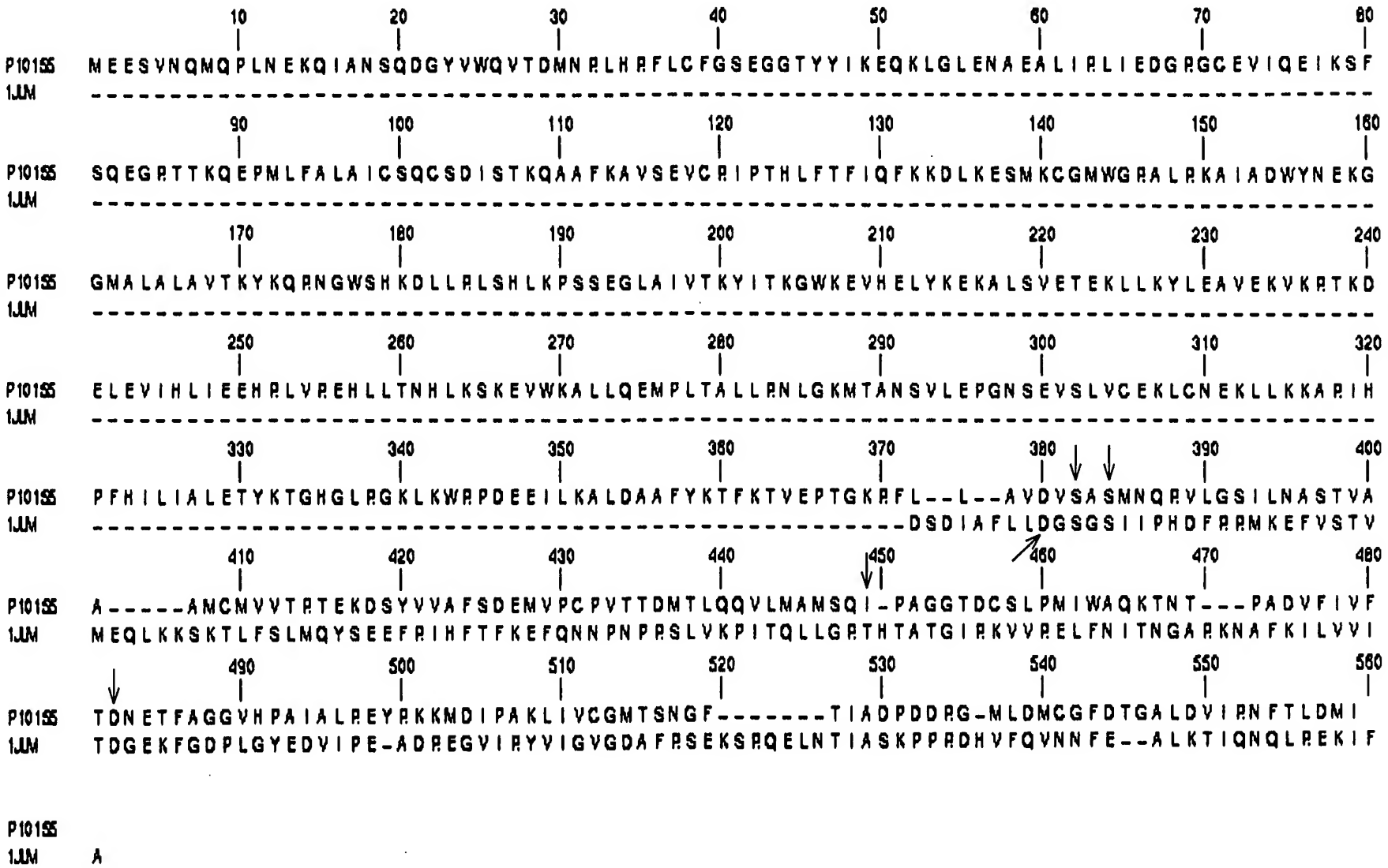
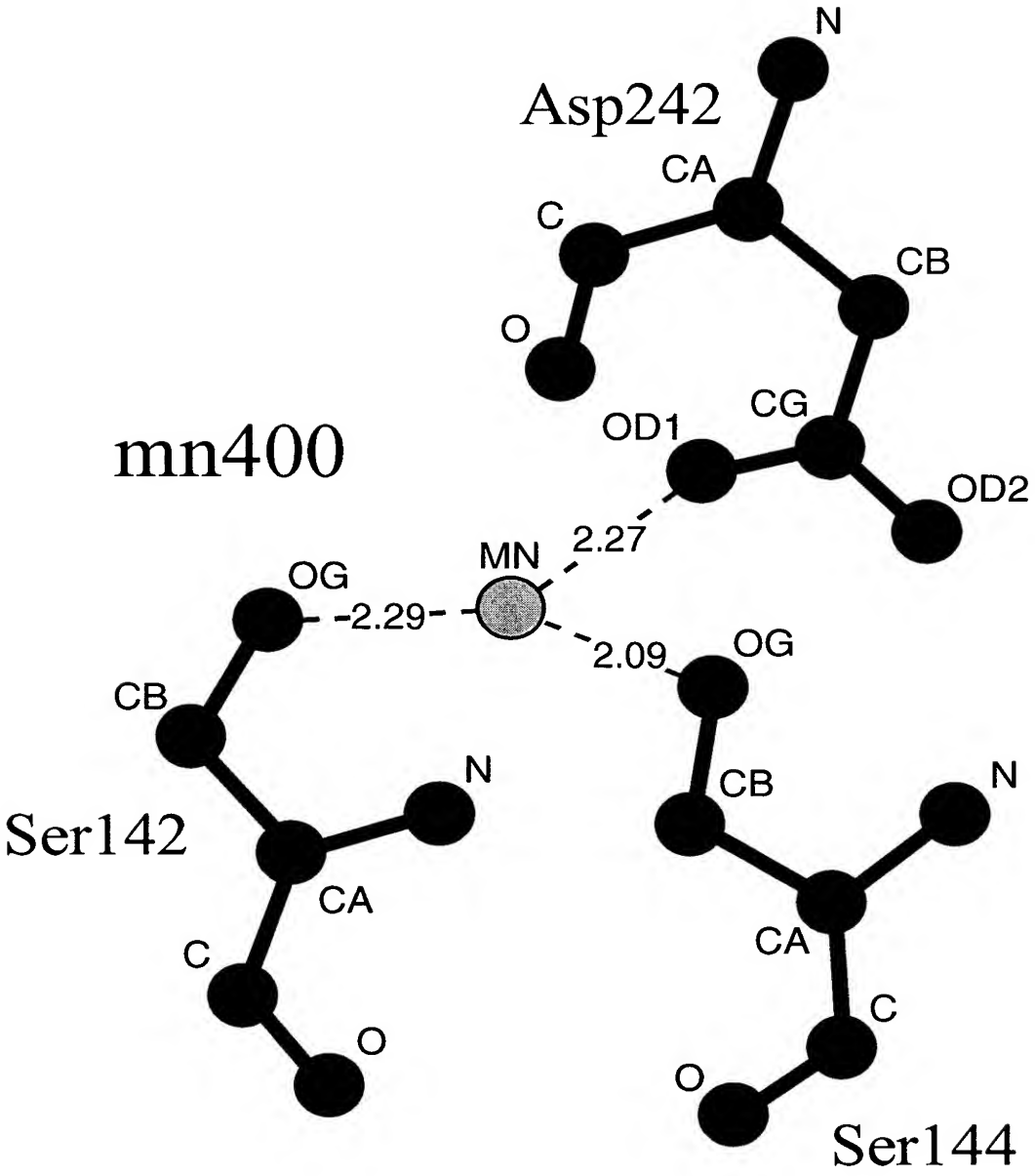
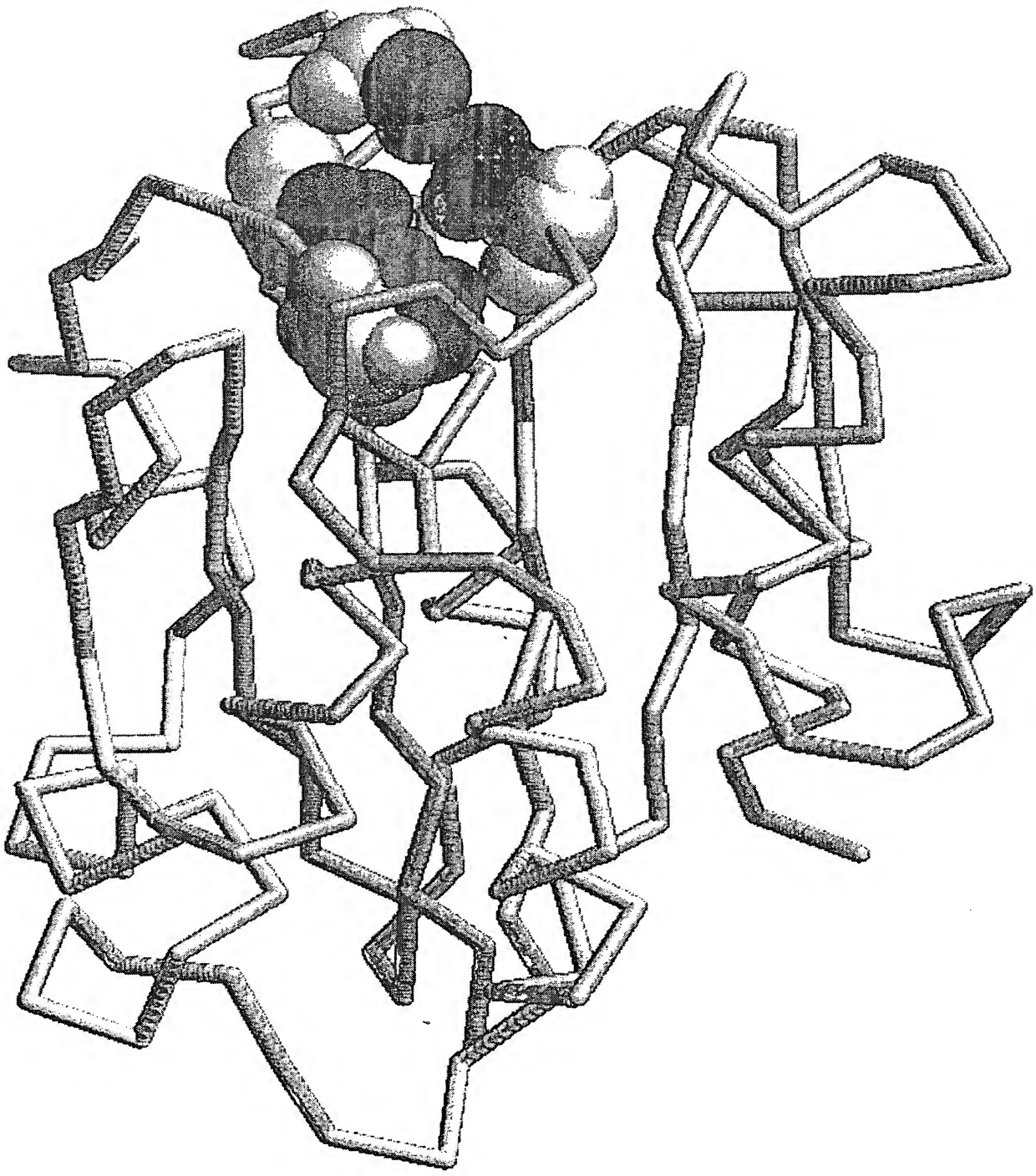


FIG. 24A



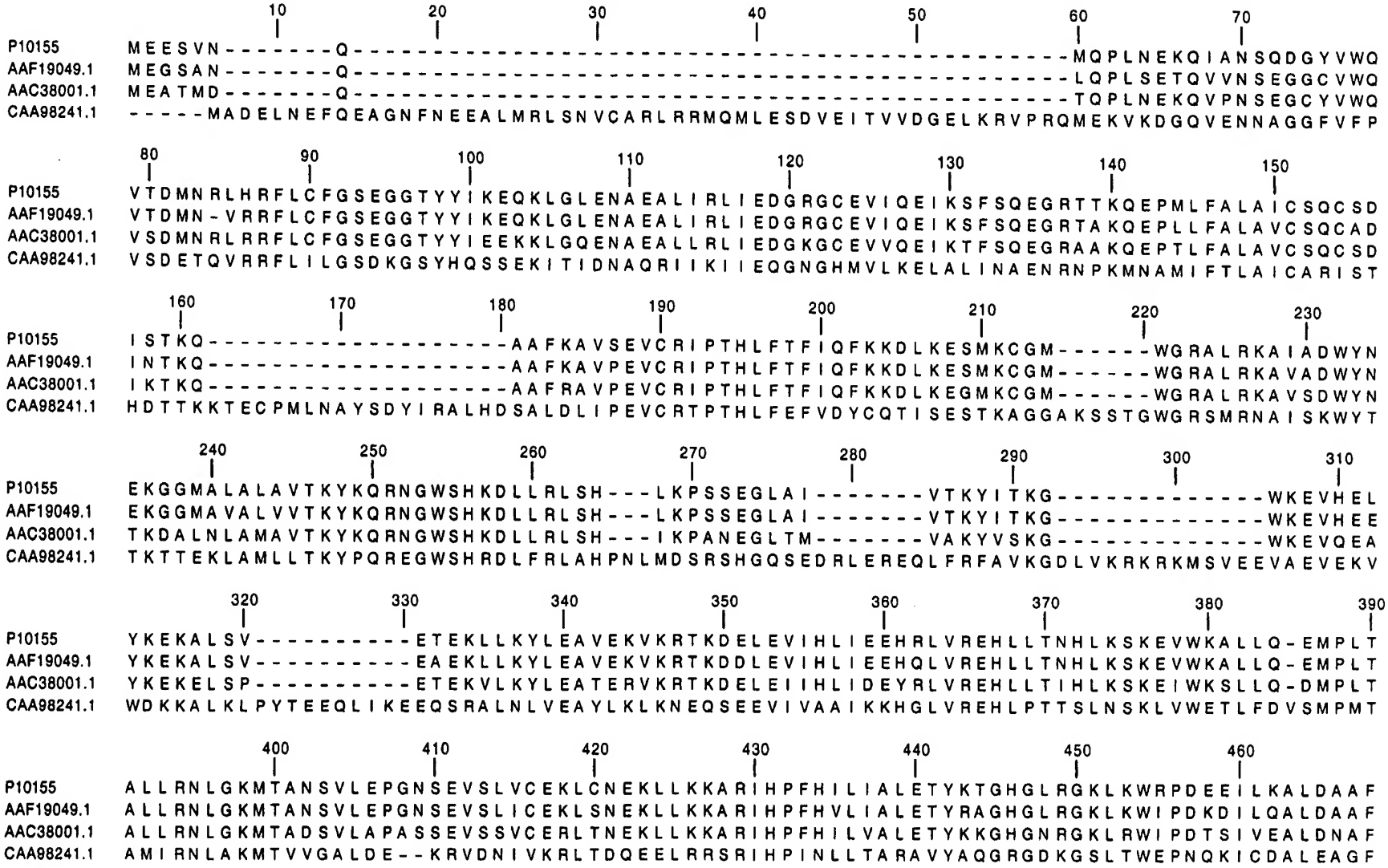

1jlm: MN 400

FIG. 24B



# FIG. 25

AlEye output (January 4, 2001 3:18 PM)



# FIG. 25 (contd.)

AlEye output (January 4, 2001 3:18 PM)